Psychological Review

# Exploring the Conceptual Universe

Charles Kemp

Online First Publication, August 27, 2012. doi: 10.1037/a0029347

Psychological Review

0033-295X/12/$12.00 DOI: 10.1037/a0029347

# Exploring the Conceptual Universe

Charles Kemp
Carnegie Mellon University

Humans can learn to organize many kinds of domains into categories, including real-world domains such as kinsfolk and synthetic domains such as sets of geometric figures that vary along several dimensions. Psychologists have studied many individual domains in detail, but there have been few attempts to characterize or explore the full space of possibilities. This article provides a formal characterization that takes objects, features, and relations as primitives and specifies conceptual domains by combining these primitives in different ways. Explaining how humans are able to learn concepts within all of these domains is a challenge for computational models, but I argue that this challenge can be met by models that rely on a compositional representation language such as predicate logic. The article presents such a model and demonstrates that it accounts well for human concept learning across 11 different domains.

*Keywords:* concept learning, compositionality, conceptual complexity, minimum description length, predicate logic

*Supplemental materials:* http://dx.doi.org/10.1037/a0029347.supp

Humans think about a wide range of concepts, including some which correspond to single words (e.g., bicycle, brown, and brother) and some which do not (e.g., brown bicycles owned by my brother). Concepts can be viewed in a variety of ways, but one common approach treats a concept as a function that picks out a category, or a subset of the items in a given domain. For example, the domain of artifacts includes many different items, and the concept of "bicycle" picks out a category that includes a subset of the items in the domain. Psychologists have studied a number of real-world domains, including artifacts, colors, and kinsfolk, and have also worked with simple synthetic domains in an attempt to understand the basic principles that govern the acquisition and use of concepts. Table 1 shows some of these domains, but includes only a tiny fraction of the *conceptual universe*, or the full space of possibilities.

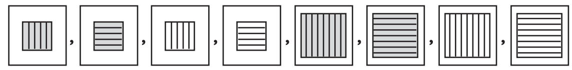
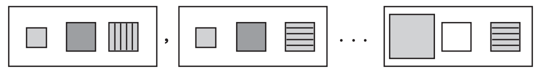
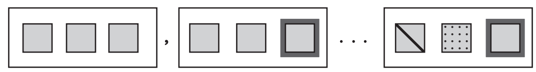
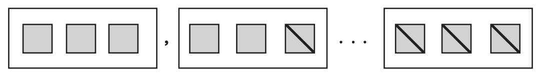
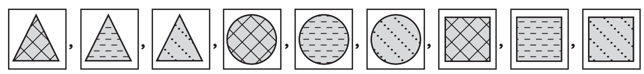
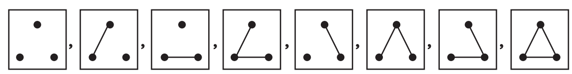
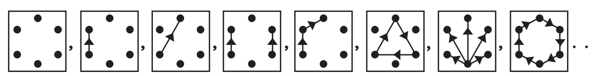
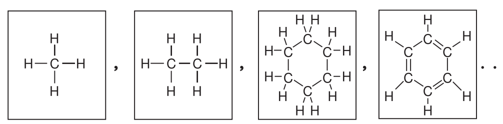
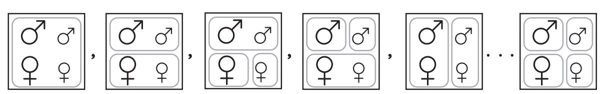
The notion of the conceptual universe motivates the three goals of the present article. The first goal is to provide a systematic account of the domains that belong to the conceptual universe and to characterize the qualitatively different categories that exist within these domains. The second goal is to describe novel and existing empirical studies that explore concept learning across multiple domains in the conceptual universe. The third goal is to develop and evaluate a computational approach to concept learning that can be applied across all of the domains in the conceptual universe. I now expand on each goal in turn.

In order to understand how humans are able to think about all of the domains in the conceptual universe, the first step is to characterize the structure of the universe. A successful characterization should allow researchers to understand the relationships between different kinds of learning problems and to identify useful targets for empirical studies. This article provides a characterization that treats objects, features, and relations as basic conceptual elements, and specifies conceptual domains by combining these elements in different ways. For example, domains 1 and 2 in Table 1 differ according to whether the features belonging to each item are distributed across several objects (domain 2) or possessed by a single object (domain 1; Shepard, Hovland, & Jenkins, 1961). Domains 2 and 3 differ according to whether the features are *substitutive* features that take two values (e.g., size is either big or small in domain 2) or *additive* features that can be described as present or absent (e.g., the slash in domain 3; Garner, 1978a; Gati & Tversky, 1982). Domains 2–4 illustrate the role of composition, because each item in these domains is constructed by combining three objects. Additional examples of composition are provided by domains 6–9, which include items such as molecules and kinship systems that correspond to configurations of objects, features, and relations. Previous researchers have characterized different kinds of objects, features, and relations (Aitkin & Feldman, 2006; Cottrell, 1975; Crockett, 1982; Feldman, 2000; Garner, 1978a; Lee & Navarro, 2002; Shepard et al., 1961; Tversky, 1977) and have emphasized the idea that these elements can be compositionally combined (Bourne, 1970; Goodman, Tenenbaum, Feldman, & Griffiths, 2008). The characterization of the conceptual universe

Code and data are available at www.charleskemp.com/code/conceptualuniverse.html. An early version of this work was presented at the 23rd Annual Conference on Neural Information Processing Systems, December 2009, Vancouver, British Columbia, Canada. This work was supported in part by National Science Foundation Grant CDI-0835797 and by the Pittsburgh Life Sciences Greenhouse Opportunity Fund. Many of the ideas in this article grew out of previous collaborations with Noah Goodman and Joshua Tenenbaum. I thank Faye Han, Rukmani Sahay, and Maureen Satyshur for assisting with the experiments, and Michael Lee, Mike Oaksford, and Ronaldo Vigo for valuable contributions.

Correspondence concerning this article should be addressed to Charles Kemp, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Baker Hall 340T, Pittsburgh, PA 15213. E-mail: ckemp@cmu.edu

Table 1
*Nine Domains That Belong to the Conceptual Universe*

| Item description | Domain |
|---|---|
| 1. 1 object, 3 substitutive binary features |  |
| 2. 3 objects, 3 substitutive binary features |  |
| 3. 3 objects, 3 additive binary features |  |
| 4. 3 objects, 1 additive binary feature |  |
| 5. 1 object, 2 substitutive ternary features |  |
| 6. 3 objects, 1 undirected binary relation |  |
| 7. 6 objects, 1 directed binary relation |  |
| 8. Multiple objects, multiple features and relations |  |
| 9. 4 objects, multiple features and relations |  |

*Note.* A domain is a set of items, and each item includes some number of objects and some number of features and relations defined over these objects. For example, each item in domain 1 includes a single object that has three features: size, shading, and stripe orientation. Each item in domain 8 is a molecule constructed by combining atoms of carbon and hydrogen. Each item in domain 9 includes four siblings of a focal individual who is not shown. The features and relations defined over these siblings include a sex feature and a relative age relation that indicates whether each sibling is older or younger than the focal individual. The gray ovals indicate different ways in which these siblings can be organized into categories.

developed here builds on the insights of these researchers, and the goal here is to systematize and extend previous efforts.

After characterizing the conceptual universe, the next step is to study human learning across multiple domains. Studying multiple domains is important because different domains expose different aspects of conceptual structure. For example, domain 4 in Table 1 can be used to explore concepts that rely on quantification, including the concept that picks out all items where two or more objects have slashes. No corresponding concept can be formulated in domain 2. Previous researchers have emphasized the importance of studying multiple domains. For example, Shepard et al. (1961), Goodwin (2006), and Mathy (2010) have compared learning patterns across multiple domains, and Feldman (2000) has explored two versions of domain 1 in Table 1 that include different numbers of binary features. Here, I discuss empirical results for several domains that have previously been studied, and present results for several novel domains, including domains that highlight the role of composition, quantification, and relations.

There are many computational accounts of concept learning that make a variety of theoretical commitments. For example, one class of models proposes that concepts correspond to mental rules (Bourne, 1970; Bruner, Goodnow, & Austin, 1956; Feldman, 2006; Fific, Little, & Nosofsky, 2010; Goodman et al., 2008; Lafond, Lacouture, & Cohen, 2009), and another proposes that concepts are better described as patterns of weights within connectionist networks (Kruschke, 1992; Love, Medin, & Gureckis, 2004). The first two contributions of this article should be relevant to researchers from both camps. Characterizing the conceptual universe and collecting empirical data help to chart the phenomena that both groups of researchers should ultimately aim to explain. The third goal of this article, however, is to develop a specific modeling approach that can account for concept learning across multiple domains. Computational theories of concept learning often focus on a single domain, but characterizing the conceptual universe motivates the development of theories that apply to all of the domains in the universe.

The first challenge is to characterize the mental representations that support concept learning. It may ultimately be possible to develop connectionist models that account for learning across all of the domains in the conceptual universe, but the path toward this goal is far from clear. In contrast, I suggest that a rule-based approach provides a natural account of learning across multiple domains. Like previous rule-based models, the model developed here assumes that humans are equipped with representational elements that pick out objects, features, and relations, and can combine these elements to construct representations of complex concepts. Importantly, these basic conceptual elements can be assembled in different ways to capture the structure of different domains. To turn this general idea into a fully-specified computational model, it is necessary to characterize the compositional language that is used to construct complex representations out of simpler parts. Predicate logic is the most familiar language of this kind, and has previously served as the representational foundation for many psychological models, including accounts of knowledge representation (Hayes, 1978; Hobbs & Moore, 1985), deductive reasoning (Braine & O'Brien, 1998; Rips, 1994; Stenning & van Lambalgen, 2008), and analogical inference (Gentner, 1983; Holyoak & Thagard, 1989). Here, I explore the extent to which predicate logic can help to account for human learning across multiple domains in the conceptual universe.

The hypothesis that concepts are represented in a compositional language immediately provides a complexity ordering over the resulting space of concepts, where the complexity of any concept corresponds to the length of its minimal description in the representation language (Chater, 1999; Chater & Vitanyi, 2003b; Fass & Feldman, 2003; Feldman, 2000). Given the assumption that the difficulty of learning a concept is predicted by its complexity, it follows that any concrete proposal about the language of mental representation will make predictions about the relative difficulty of learning different concepts. Here, I propose that description length in predicate logic can predict the difficulty of human learning across multiple domains in the conceptual universe. Although previous studies have not directly explored the relationship between conceptual complexity and description length in predicate logic, there are previous accounts of conceptual complexity that focus on propositional or Boolean logic. The literature on Boolean concept learning is extensive (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Shepard et al., 1961), and Feldman (2000; see also Neisser & Weene, 1962) has reported that the psychological complexities of Boolean concepts are well predicted by their description lengths in propositional logic. The present article builds on Feldman's work in two key respects. First, I focus on concepts that rely on quantification and relations and therefore probe aspects of mental representation that go beyond propositional logic. Second, I show how theories of mental representation can be informed by comparing concept learning across qualitatively different domains.

The next section provides a formal characterization of the conceptual universe and describes a method for identifying the different types of concepts that exist within each domain. I then describe a rule-based account of concept learning that relies on predicate logic and use it to account for data collected across 11 qualitatively different domains. The analyses consider data from two new experiments along with data from experiments carried out by several previous researchers (Aitkin & Feldman, 2006; Crockett, 1982; Feldman, 2000; Kemp, Goodman, & Tenenbaum, 2008a). The results suggest that a rule-based approach that relies on predicate logic as a representation language is capable of accounting for concept learning across a large part of the conceptual universe.

## The Conceptual Universe

Before discussing the experimental study of concept learning, it is important to consider the range of settings within which concept learning can occur. Humans can learn and think about many kinds of concepts, and characterizing the full range of possibilities establishes the broader framework within which specific empirical investigations should be situated. This article formalizes concepts as functions that pick out a subset of the items in a domain, and characterizing the space of possible concepts therefore requires a characterization of the space of possible domains. Nine domains are shown in Table 1, and the full set of domains is referred to as the conceptual universe.

This section provides a characterization of the conceptual universe that is far from exhaustive, but that nevertheless provides a useful foundation for experimental and computational work. Each domain in the universe is a collection of items, and specifying a space of possible items therefore provides a way to characterize a broad family of domains. Each item is formalized here as a *semantic system* of objects, features, and relations. Figure 1 shows examples of four semantic systems and includes set-theoretic specifications of each one. Figure 1a is a simple system $S$ that includes three objects and one binary feature. The type specification takes the following form:

$$S = (O = \{o_1, o_2, o_3\}, F : O \rightarrow \{0, v_1\}).$$

The specification refers to a set $O$ of three objects $o_1$, $o_2$, and $o_3$ that correspond to the three gray squares in Figure 1a. The binary feature $F$ takes two values that can be described as 0 (absent) or $v_1$ (slash present). For example, suppose that the system in Figure 1a is a card issued by a dining hall where the three squares represent three stations that serve appetizers, entrees, and desserts. When a patron visits a station, a slash is added to the corresponding square on her card.

Figure 1b shows a second example that includes a binary relation in addition to a binary feature. The system is a molecule of methane and is constructed from five objects $o_1$–$o_5$ that correspond to single atoms. The feature $E$ indicates whether each atom is a carbon atom or a hydrogen atom, and the relation $Bond(x, y)$ indicates whether objects $x$ and $y$ are joined by a single bond, a double bond, or no bond.

Figure 1c shows a third example that corresponds to a simple kinship system. The system shown includes a focal individual labeled as Ego and four siblings of this individual: an elder brother (Be), younger brother (By), elder sister (Ze), and younger sister (Zy). The system includes a feature that indicates whether each individual is male or female, and a relation $Rel\_Age$ that indicates the relative ages of each pair of individuals. More extensive kinship systems may include many additional relations, including relations that pick out pairs $(x, y)$ where $x$ is the parent of $y$, and pairs where $x$ is the spouse of $y$.

Figures 1a–1c illustrate how objects, features, and relations can be combined to construct semantic systems. These systems in turn can be treated as "compound objects" over which features and relations are defined to create higher-level semantic systems. Figure 1d shows a high-level system that includes eight systems $s_1$–$s_8$,

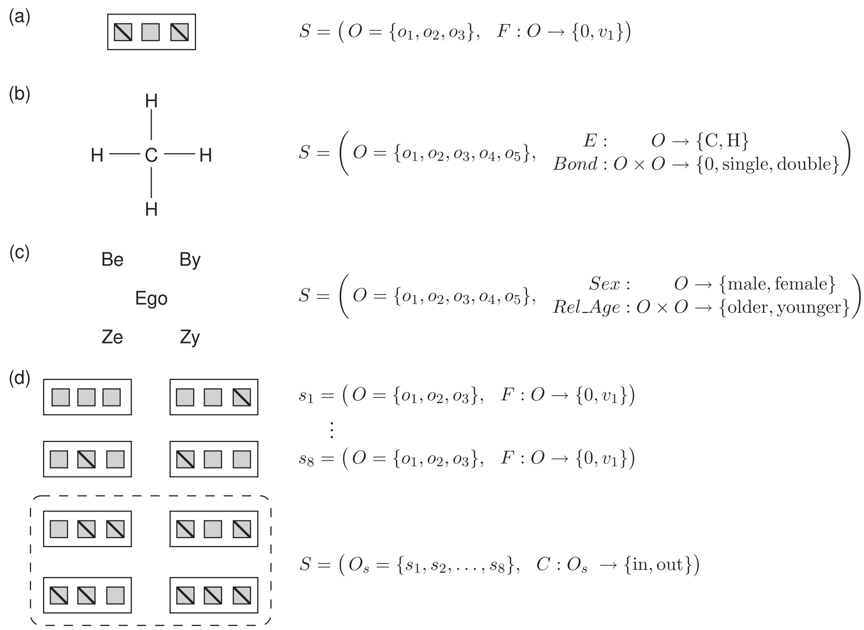


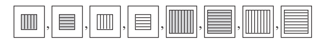
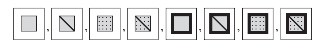
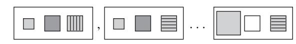
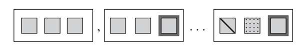
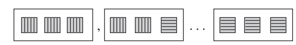
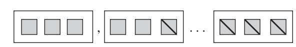
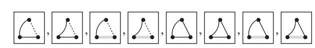
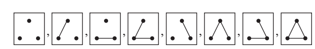

*Figure 1.* Semantic systems. (a) A system with three objects (gray squares) and a single binary feature (the slash). (b) A molecule of methane can be viewed as a system that includes five objects (atoms), a feature that characterizes the chemical identity of each atom, and a relation that indicates whether a single bond, a double bond, or no chemical bond exists between each pair of atoms. (c) A kinship system that includes a focal individual labeled as Ego and four siblings of this individual. The system includes a feature that specifies the sex of each individual, and a relation that indicates the relative age of each pair of individuals. (d) A high-level semantic system *S* defined over eight compound objects $s_1$–$s_8$, each of which is a semantic system in its own right. The system shown includes a feature *C* that picks out a category that includes all items with two or more slashes.

each of which has the type specification shown in Figure 1a. The higher-level system includes a feature *C* defined over these systems that picks out all systems with two or more slashes. For example, suppose that patrons of the dining hall are eligible for a free drink if they make purchases at two or more stations. Category *C* includes all cards that qualify the bearer for a free drink. Similar examples can be formulated within the chemical domain shown in Figure 1b. Consider, for example, the category of "aromatic hydrocarbons," which includes molecules that have a ring of carbon atoms with alternating double and single bonds. This category can be formalized using a feature that picks out all qualifying molecules from the family shown in domain 9 of Table 1. The process of defining higher-level features and relations can continue at progressively higher levels. For example, chemical elements are combined to form molecules, which are combined to form cells, which are combined to form organisms, and scientists are able to think about features and relations at all of these levels.

Although Figure 1 focuses on some of the simplest possible cases where objects, features, and relations are combined to construct semantic systems, the same kind of compositional structure is characteristic of everyday thought. For example, collective nouns such as "family" and "committee" refer to groups of objects that are mentally bound together into systems (Bloom & Kelemen, 1995). Events can also be viewed as systems of objects and relations. For example, an event where Mary gives a book to John can be characterized as a semantic system that specifies a relationship involving three objects: Mary, the book, and John (Davidson, 1967; Fillmore, 1968; Jackendoff, 1983; Levin & Rappaport Hovav, 1995; Pustejovsky, 1991). Similarly a robbery event can be viewed as a semantic system that specifies relationships between objects that include the thief, the victim, and the goods that were stolen (Gentner & Kurtz, 2005). Higher-level systems can be formulated in turn by defining features or relations over events. For example, the category of "armed robberies" can be formalized as a higher-level system which includes a higher-level feature that picks out all robbery events where a weapon was used to carry out the crime. Similarly, higher-level relations can be used to organize events into systems that correspond to stories (Rumelhart, 1975), narratives, or scripts (Schank & Abelson, 1977). Examples of this kind suggest that humans readily organize objects into semantic systems, and find it natural to think about features and relations defined over these systems. The need to account for this kind of compositional structure is a key motivation for the model of concept learning developed in later sections.

The set-theoretic specifications in Figure 1 were introduced relatively informally, but the space of possible specifications can be given a formal characterization. The basic idea is to characterize the basic conceptual elements that appear in these specifications, then to specify a compositional process that allows these elements to be combined. Generating possible specifications in this way characterizes a vast space of domains that can be used to explore human concept learning. Table 2 shows a corner of this space that is organized around a classic domain studied by Shepard et al. (1961), where each item includes a single object with three binary

Table 2
*Specifications for Eight Domains That Include Eight Items Each*

| Domain label | Domain specification | Domain | # types |
|---|---|---|---|
| 1. (1O, 3SF) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1\}, \begin{array}{l} F_1 : O \to \{v_1, v_2\} \\ F_2 : O \to \{v_3, v_4\} \\ F_3 : O \to \{v_5, v_6\} \end{array} \right)$ |  | 6 |
| 2. (1O, 3AF) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1\}, \begin{array}{l} F_1 : O \to \{0, v_1\} \\ F_2 : O \to \{0, v_2\} \\ F_3 : O \to \{0, v_3\} \end{array} \right)$ |  | 10 |
| 3. (3O, 3SF) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1, o_2, o_3\}, \begin{array}{l} F_1 : o_1 \to \{v_1, v_2\} \\ F_2 : o_2 \to \{v_3, v_4\} \\ F_3 : o_3 \to \{v_5, v_6\} \end{array} \right)$ |  | 6 |
| 4. (3O, 3AF) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1, o_2, o_3\}, \begin{array}{l} F_1 : o_1 \to \{0, v_1\} \\ F_2 : o_2 \to \{0, v_2\} \\ F_3 : o_3 \to \{0, v_3\} \end{array} \right)$ |  | 10 |
| 5. (3O, 1SF) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1, o_2, o_3\}, \; F : O \to \{v_1, v_2\} \right)$ |  | 9 |
| 6. (3O, 1AF) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1, o_2, o_3\}, \; F : O \to \{0, v_1\} \right)$ |  | 10 |
| 7. (3O, 3SR) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1, o_2, o_3\}, \begin{array}{l} R_1 : \{\{o_1, o_2\}\} \to \{v_1, v_2\} \\ R_2 : \{\{o_1, o_3\}\} \to \{v_3, v_4\} \\ R_3 : \{\{o_2, o_3\}\} \to \{v_5, v_6\} \end{array} \right)$ |  | 6 |
| 8. (3O,1AR) | $\{S_1, \ldots, S_8\}$ where $S_i = \left( O = \{o_1, o_2, o_3\}, \; R : \{\{o_i, o_j\}\} \to \{0, v_1\} \right)$ |  | 10 |

*Note.* Each item in the first two domains includes a single object (1O), and the items in all remaining domains include three objects (3O) each. Each domain includes either substitutive features (SF), additive features (AF), a substitutive relation (SR), or an additive relation (AR). The final column shows the number of four-item category types that can be formulated within each domain. All domains can be viewed as variants of the domain originally studied by Shepard et al. (1961). Many of these domains have been studied by previous researchers: (1O,3SF) (Feldman, 2000; Mathy & Bradmetz, 2004; Nosofsky, Gluck, et al., 1994; Shepard et al., 1961; Vigo, 2009); (1O,3AF) (Mathy & Bradmetz, 2011; Sakamoto & Love, 2004); (3O,3SF) (Shepard et al., 1961); (3O,1SF) (Shepard et al., 1961); (3O,1AF) (Goodwin, 2006; Mathy, 2010); and (3O,1AR) (Crockett, 1982).

features. This classic domain appears in the first row of Table 2, and the remaining domains in Table 2 show variants of the classic domain that are created by combining objects, features, and relations in different ways. Characterizing and exploring these different domains is important because we will see that different domains in Table 2 can lead to different patterns of concept learning.

The domains in Table 2 include both *additive* and *substitutive* features (Gati & Tversky, 1984; see also Garner, 1978a, 1978b, for a similar distinction). Additive features such as the slash in domain 6 can be described as either absent or present. A slash is a relatively abstract example of an additive feature, but everyday examples of additive features include binary features that specify whether a person wears glasses and has a moustache (Gati & Tversky, 1984). The domain specifications in Table 2 represent additive features as features that can take a value of 0 (absent). Substitutive binary features such as the texture feature in domain 5 also take one of two values, but these two values are symmetric—there is no sense in which one feature value (e.g., horizontal) corresponds to the presence of the feature and the other (e.g., vertical) corresponds to the absence of the feature. Domains 7 and 8 in Table 2 show that relations can also be substitutive or additive. To keep track of the different kinds of relations and features, I refer to the eight domains using the labels in the second column of Table 2. For example, the first domain includes systems that are constructed using one object and three substitutive features, and is therefore referred to as domain (1O,3SF). The second domain is similar but based on additive rather than substitutive features, and is therefore referred to as domain (1O,3AF).

Future researchers may find ways to improve on the set-theoretic notation used here to characterize the conceptual universe. More important than the formal machinery, however, are the foundational ideas that motivate the machinery. The approach described in this section is founded on two key ideas: first, that objects, features, and relations are the basic elements that can be combined to construct descriptions of the world, and second, that these conceptual elements can be bound together into compound objects or systems. The same basic ideas lie at the heart of many approaches to knowledge representation, and the contribution here is to apply these familiar ideas in a way that supports the study of human concept learning.

There are at least two reasons why characterizing the conceptual universe may be productive. First, comparing learning phenomena across closely-related domains can provide a powerful tool for understanding how humans learn and represent concepts. For example, Experiment 1 in this article compares learning times across four of the domains in Table 2 in an attempt to explore how quantification shapes concept learning. Similar approaches have been productive in other areas of psychology—for example, problem-solving researchers have found it useful to explore why isomorphic problems can lead to very different patterns of behavior (K. Kotovsky, Hayes, & Simon, 1985). Second, understanding the scope of the universe can help to establish the most important priorities for psychological research. Current studies often aim to provide a detailed account of phenomena within a single domain, but across-domain coverage is arguably just as important as within-domain coverage, if not more so (Lee, 2011; Newell, 1989). Previous accounts of conceptual complexity have focused on selected domains from Table 1, but no existing account can explain how humans learn concepts across all of these domains. The formal approach described in later sections is motivated in large part by the challenge of capturing learning across the entire conceptual universe.

## Concept Types

Now that the conceptual universe has been characterized, I turn to the problem of characterizing the different types of concepts that exist within each domain. Consistent with Figure 1d, a concept will be formalized as a high-level feature that picks out a subset of the items in a domain. The subset picked out in this way is called the *extension* of a concept. Appendix A describes a general method for identifying the number of qualitatively different extensions that exist within any domain. This section describes the number of four-item extensions that exist within the domains in Table 2. Focusing on these domains will illustrate that domains which appear similar on the surface may support different kinds of concepts, and will help to motivate the experiments that follow.

All of the domains in Table 2 include exactly 8 items. As a result, each domain allows $\binom{8}{4} = 70$ different ways to specify a category extension that includes exactly four items. Some of these extensions have the same basic structure. In domain (1O,3SF), for example, the extension that includes only the four small items and the extension that includes only the four gray items are structurally similar, because both can be described using a single feature value. Both extensions can therefore be treated as instances of the same type. Based on considerations of this kind, Shepard et al. (1961) pointed out that the 70 concepts of size four in domain (1O,3SF) can be organized into six qualitatively different types, and I refer to these types as SHJ types I–VI. A representative of each type is shown in Figure 2b. The eight vertices of the cube in Figure 2a represent the eight items in the domain, and the labels of these vertices were created by converting the three features to Boolean values. The shaded nodes in Figure 2b show items that belong to the extension of a concept. For example, the representative of type I includes all nodes where the second feature has value 1, and corresponds to the extension that includes the four items in domain (1O,3SF) where feature $F_2$ takes value $v_4$.

Appendix A describes a generalization of the analysis given by Shepard et al. (1961) that can be used to identify the concept types that exist within the domains in Table 2. The number of four-item types for each domain is shown in the final column of Table 2. All of these domains include eight items, and each item in these domains can be represented as a triple of Boolean values. For example, the undirected relations in domain (3O,1AR) have three possible edges, and the three Boolean variables in this case indicate whether each edge is present or absent. It is tempting to assume that the six SHJ types apply to each domain, and Appendix A mentions several groups of researchers (including Shepard et al., 1961) who appear to have made this assumption. It turns out, however, that the six SHJ types apply to only three of the domains.

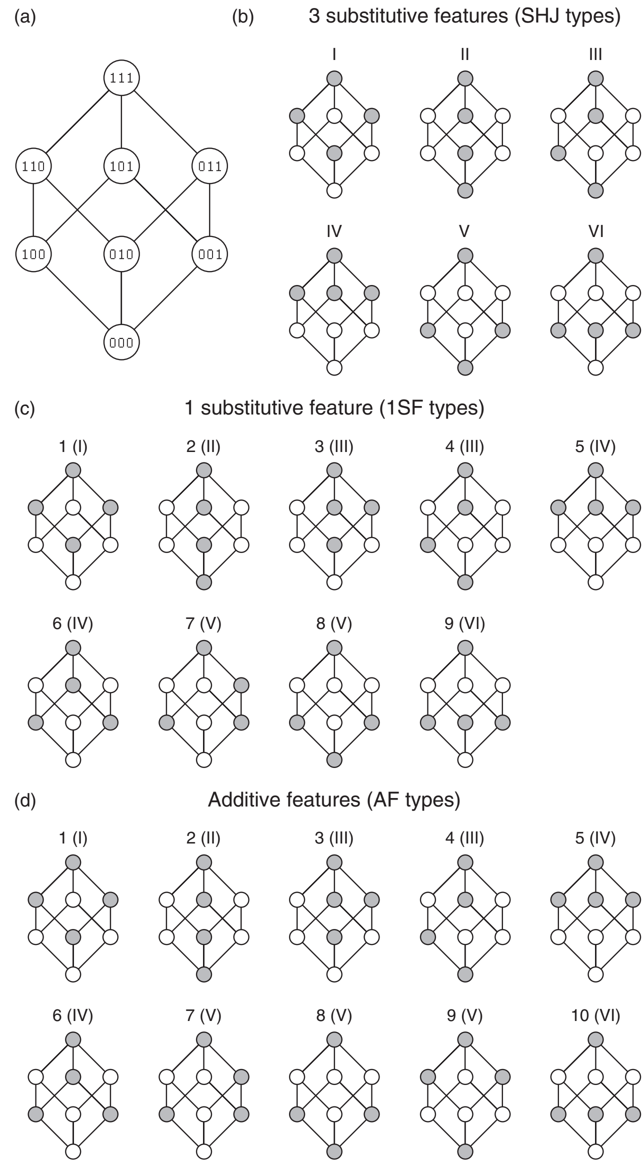



*Figure 2.* Concept types for the domains in Table 2. (a) A stimulus lattice for domains (including all domains in Table 2) where each item can be encoded as a triple of binary values. (b) Domains like (1O,3SF) and (3O,3SF) include six types of size four that will be called the six SHJ types. A single concept representing each type is shown, where gray nodes indicate items that belong to the extension of the concept. (c) Domains like (3O,1SF) include nine types of size four. The SHJ type for each concept is shown in parentheses. (d) Domains like (3O,1AF) include 10 types of size four.

Domain (3O,1SF) has the nine types shown in Figure 2c, and domains (1O,3AF), (3O,3AF), (3O,1AF), and (3O,1AR) have the 10 types shown in Figure 2d. To see why these domains produce different numbers of types, it is helpful to consider types 5 and 6 in Figure 2d. Both types correspond to SHJ type IV, but the two are qualitatively different within the context of domain (3O,1AF). Type 5 can be described as a category that includes all items where two or more objects take value $v_2$ on feature $F$, but Type 6 has no similar description. Experiment 1 in this article provides some empirical evidence that humans are sensitive to distinctions of this kind that are not captured by the six SHJ types.

## A Rule-Based Account of Concept Learning

A long-term challenge for psychological research is to explain how concept learning operates across all the domains in the conceptual universe. The formal analyses described in previous sections support progress toward this goal by characterizing a large collection of domains and identifying the types of concepts that exist within these domains. These analyses are intended to be theoretically neutral, and should therefore be relevant to researchers from many different theoretical persuasions. The goal so far has been to characterize the many kinds of concepts that exist, not to make any psychological claims about human learning or mental representation.

The remaining sections of the article focus on psychological data and computational models of learning. When characterizing the conceptual universe, it was convenient to treat a concept as a high-level system

$$(O_s = \{s_{1,} s_{2,} \ldots, s_n\}, C : O_s \rightarrow \{\text{in, out}\}),$$

where $\{s_1, \ldots, s_n\}$ is the set of items in a given domain, and $C(s_i) =$ in if and only if item $s_i$ belongs to the extension of the concept. This extensional view of concepts is useful for some purposes, but in order to understand how humans learn and think about concepts it is critical to consider the *intension* of the function $C(\cdot)$. The intension is the mental representation that allows a learner to decide whether any given item is an instance of the concept. For example, the intension might correspond to a rule, a prototype, or a set of memorized exemplars.

A challenge for computational modelers is to describe some way of constructing intensions that applies across all of the domains in the conceptual universe. The rest of the article explores the idea that these intensions are constructed in a compositional representation language. The language treats objects, features, and relations as basic elements, and provides a way to combine these elements in order to construct the intension of a concept. The compositional nature of the language allows a relatively small set of basic elements to combine in many different ways that are needed to capture qualitatively different concepts across qualitatively different domains. The predictions of this approach depend critically on the specific language chosen, and the language proposed here is characterized in the next section. After introducing this language, I discuss how the resulting model is related to previous models of concept learning.

The hypothesis that mental representations are constructed from a compositional language can help to explain both knowledge representation and learning (Fodor, 1975). Generating concepts from a compositional language automatically provides a complexity ordering over the resulting space of concepts, where the complexity of any concept corresponds to the length of its minimal description in the representation language (Chater & Vitanyi, 2003b; Fass & Feldman, 2003; Feldman, 2000; Kemp et al., 2008a; Kemp & Jern, 2009a). Following previous researchers, I explore the hypothesis that the minimal description length of a concept determines its subjective complexity, or the ease with which it is learned. Later sections of the article refer to this hypothesis as the *description length hypothesis*.

In order for learners to identify the minimal description of a concept they must use some procedure to search through the space of descriptions. A complete account of human learning will need to characterize the search process in detail, but the description length hypothesis is appealing in part because it abstracts away these details. As a result, the hypothesis provides a simple initial strategy for evaluating claims about mental representation. For example, the hypothesis can be used to generate parameter-free predictions for any proposed representation language, and predictions of this sort will be used in this article to evaluate the relative merits of several representation languages including propositional logic and several varieties of predicate logic. Comparisons of this sort can help to identify the representational assumptions that best account for human learning, and subsequent research can then aim to account for the data even more closely by combining these representational assumptions with additional assumptions about processing.

Subjective complexity can be operationalized in a variety of ways, and experimental paradigms that focus on learning, memory, and inductive reasoning are all relevant. A concept is subjectively simple to the extent that it is rapidly learned and accurately remembered. When participants do make errors in learning or recalling a concept, their errors will tend to be consistent with concepts that are subjectively simpler than the true underlying concept. Similarly, if participants are provided with incomplete information about a concept then asked to generalize the concept to novel items, their inferences will tend to be consistent with a subjectively simple concept that is compatible with the data that they have seen. Although different measures of subjective complexity can sometimes lead to qualitatively different results, previous researchers have found that these measures often converge on a single, stable characterization of the relative complexities of the concepts within a domain (Shepard et al., 1961). Later sections of this article use learning time and classification accuracy as measures of subjective complexity.

### Predicate Logic as a Representation Language

The representation language described in this section is motivated by two basic constraints. The first is that the language must support statements about objects, features, and relations. The assumption throughout is that objects, features, and relations are the basic conceptual units that allow humans to think about multiple domains in the conceptual universe. The second constraint is that the language must support quantification. For example, the language must provide a way to express the fact that the rightmost item in domain 4 in Table 1 is an item where *all* of the squares have a slash.

The best-known example of a representation language that satisfies both constraints is predicate logic. Other representation schemes could be considered, including frames (Barsalou, 1992; Minsky, 1975) and semantic networks (Shapiro & Rapaport, 1992; Sowa, 1984), but here I explore the extent to which predicate logic can account for concept learning across multiple domains. There are many versions of predicate logic, including first-order logic, higher-order logics, and intensional logics (Montague, 1973). Here, I focus on a language that is closely related to standard first-order logic. First-order logic is at best a rough approximation of the conceptual resources that support concept learning (Jackendoff, 1983; Thomas, 2009), but working with a well-known language allows a simple initial investigation of the description length hypothesis. To the extent that this initial investigation is successful, future research can explore whether alternative representation languages provide a more accurate account of human learning.

The first-order language explored here is referred to as language OQ, and is compared to two alternatives called FQ and OQ + FQ. Language OQ supports quantification over objects, language FQ supports quantification over features, and language OQ + FQ supports quantification over objects and features. A grammar for generating rules in language OQ + FQ is shown in Figure 3, and five such rules are shown in the second column of Table 3. The first production in Figure 3 indicates that each rule takes the form $\forall i C(i) \leftrightarrow$ disjunction, where the right hand side specifies conditions that must be satisfied if item i belongs to concept *C*. The right hand side is an expression in disjunctive normal form: that is, a disjunction of conjunctions, where each conjunction specifies a sufficient condition for belonging to concept *C*. Note that $\vee$ is a symbol for OR, and $\wedge$ is a symbol for AND. Each conjunction is built from *literals,* including literals like $F(a) = 0$ and $F(a) \neq 0$, which indicate that object a does or does not take a certain value for feature F. Literals like $R(a, b) = 1$ and $R(a, b) \neq 1$ indicate that a relation R does or does not hold between two objects a and b.

The resources described so far are sufficient to generate rules in propositional logic. Predicate logic, however, also supports quantification, and languages OQ, FQ, and OQ + FQ all include two quantifiers: for all ($\forall$) and there exists ($\exists$). Languages OQ and OQ + FQ allow quantification over objects. For example, both languages can express a rule which indicates that all objects have feature F ($\forall x\ F(x) = 1$, where x is a variable that ranges over all objects in the domain). Quantifiers can be nested—for example, $\forall x\ \exists y\ R(x, y) = 1$ indicates that for all objects x there is some object y such that $R(x, y) = 1$. When quantifiers are nested, different variables are assumed to refer to different objects. For example, $\forall x\ \exists y\ F(y) = 1$ indicates that for all x, there is some y other than x such that $F(y) = 1$. Languages FQ and OQ + FQ allow quantification over predicates. For example, both languages can express a rule which indicates that object a has all features under consideration ($\forall Q\ Q(a) = 1$, where Q is a variable that ranges over all features in the domain). The grammar in Figure 3 can be adjusted to specify language OQ, FQ, or OQ + FQ by editing production 10. Language OQ only allows the possibility of ovars, or variables that refer to objects. Language FQ only allows the possibility of pvars, or variables that refer to predicates (i.e., features or relations). As shown in Figure 3, language OQ + FQ includes both variables that refer to objects and variables that refer to predicates.

| | | | |
|---|---|---|---|
| 1 | rule | $\rightarrow$ | $\forall$`i` `C(i)` $\leftrightarrow$ disjunction |
| 2 | disjunction | $\rightarrow$ | disjunction $\vee$ conjunction \| conjunction \| qstring(disjunction) |
| 3 | conjunction | $\rightarrow$ | conjunction $\wedge$ literal \| literal \| qstring(conjunction) |
| 4 | literal | $\rightarrow$ | P(o)= v \| P(o)$\neq$v \| P(o,o)=v \| P(o,o)$\neq$v |
| 5 | v | $\rightarrow$ | `0` \| `1` \| `2` \| ... |
| 6 | o | $\rightarrow$ | ovar \| `a` \| `b` \| ... |
| 7 | P | $\rightarrow$ | pvar \| `F` \| `G` \| ... |
| 8 | qstring | $\rightarrow$ | qstring qpair \| qpair |
| 9 | qpair | $\rightarrow$ | $\forall$ var \| $\exists$ var |
| 10 | var | $\rightarrow$ | ovar \| pvar |
| 11 | ovar | $\rightarrow$ | `x` \| `y` \| ... |
| 12 | pvar | $\rightarrow$ | `V` \| `W` \| ... |

*Figure 3.* A grammar for generating rules in language OQ + FQ. Grammars for languages OQ and FQ are produced by editing production 10 so that a variable must be either an object variable (ovar) or a predicate variable (pvar), respectively. A grammar for propositional logic is produced by removing qstring(disjunction) and qstring(conjunction) from productions 2 and 3.

The central column of Table 3 shows five logical rules that are expressed in full, but for readability this article primarily uses the summary representations in the final column. The first summary representation indicates that $F_a$ and $F'_a$ will be used to indicate $F(a) = 1$ and $F(a) = 0$, respectively. Conjunctions are represented by concatenating the conjuncts: for example, $F(a) = 1 \wedge H(a) = 1$ is represented as $F_aH_a$. Disjunctions are represented using the + symbol: for example, $F(a) = 1 \vee H(a) = 1$ is represented as $F_a + H_a$. Finally, the opening sections of the full concept descriptions $\forall iC(i) \leftrightarrow \ldots$ are dropped, and the summary representations show only the disjunctions on the right hand sides of the full descriptions.

Table 4 shows several summary representations of rules formulated in language OQ. The rule numbers correspond to the concept types shown in Figure 2d. Rule 1 is true of any item in domain (3O,1AF) where object b has feature F—in other words, where the second square has a slash. Rule 2 is true of any item if objects a and c either both have or both do not have a slash. Rule 4 is similar in structure, and is true of any item where b and c both do not have a slash, or where a and c both have a slash. Rule 5 is true of any item where two or more objects have a slash. The rule indicates that for every x, there is another object y that has a slash. If x is chosen to be an object with a slash, the rule implies that there must be at least one additional object with a slash. Rule 6 applies to an item if all objects have a slash, or if object b does not have a slash and at least one of the remaining objects has a slash. Finally, rule 10 applies to any item that has either three objects with a slash or exactly one object with a slash. The second part of the rule is true of all items where some object y has a slash and where every other object z does not have a slash.

Given any candidate representation language, the description length hypothesis proposes that the subjective complexity of a concept is predicted by the length of the shortest rule that captures the concept. This article assumes that the complexity of each rule is a weighted sum based on the literals that it contains. One-place literals (e.g., $F_a$, $F_x$, $F'_a$, or $F'_x$) receive a weight of one, and two-place literals (e.g., $R_{ab}$ or $R'_{xy}$) receive a weight of two. Complexity values for the five rules already described are shown in Table 4. Assigning unit weight to each one-place literal means that the complexity measure reduces to Feldman's (2000) notion of complexity when applied to a rule that does not include relations or quantification. Weighting two-place literals more heavily than one-place literals is consistent with the idea that each argument of a predicate must be mentally represented: for example, representing the fact that John reads novels is assumed to be more demand-

Table 3
*Predicate Logic Representations of Concepts From Five Domains*

| Domain | Full concept description | Summary |
|---|---|---|
| 1. (3O, 3SF) | $\forall \mathtt{i}\ \mathtt{C(i)} \leftrightarrow (\mathtt{F(a)} = 0 \wedge \mathtt{G(b)} = 0) \vee (\mathtt{F(a)} = 1 \wedge \mathtt{H(c)} = 1)$ where $\mathtt{i}_{\mathtt{obj}} = \{\mathtt{a, b, c}\}$ | $\mathtt{F}'_{\mathtt{a}}\mathtt{G}'_{\mathtt{b}} + \mathtt{F}_{\mathtt{a}}\mathtt{H}_{\mathtt{c}}$ |
| 2. (3O, 1AF) | $\forall \mathtt{i}\ \mathtt{C(i)} \leftrightarrow (\forall \mathtt{x}\, \exists \mathtt{y}\, \mathtt{F(y)} = 1) \vee (\forall \mathtt{z}\, \mathtt{F(z)} = 0)$ where $\mathtt{x, y, z} \in \mathtt{i}_{\mathtt{obj}} = \{\mathtt{a, b, c}\}$ | $\forall_{\mathtt{x}}\exists_{\mathtt{y}}\mathtt{F}_{\mathtt{y}} + \forall_{\mathtt{z}}\mathtt{F}'_{\mathtt{z}}$ |
| 3. (1O, 2SF) | $\forall \mathtt{i}\ \mathtt{C(i)} \leftrightarrow (\mathtt{F(a)} \neq 2 \wedge \mathtt{G(a)} \neq 1) \vee (\mathtt{F(a)} = 2 \wedge \mathtt{H(a)} \neq 0)$ where $\mathtt{i}_{\mathtt{obj}} = \{\mathtt{a}\}$ | $\mathtt{F}'_{2}\mathtt{G}'_{1} + \mathtt{F}_{2}\mathtt{H}'_{0}$ |
| 4. (3O, 1AR) | $\forall \mathtt{i}\ \mathtt{C(i)} \leftrightarrow (\forall \mathtt{x}\, \exists \mathtt{y}\, \mathtt{R(x, y)} = 1) \vee (\forall \mathtt{z}\, \mathtt{R(a, z)} = 0)$ where $\mathtt{x, y, z} \in \mathtt{i}_{\mathtt{obj}} = \{\mathtt{a, b, c}\}$ | $\forall_{\mathtt{x}}\exists_{\mathtt{y}}\mathtt{R}_{\mathtt{xy}} + \forall_{\mathtt{z}}\mathtt{R}'_{\mathtt{az}}$ |
| 5. (3O×2O, 1AR) | $\forall \mathtt{i}\ \mathtt{C(i)} \leftrightarrow (\forall \mathtt{x}\, \exists \mathtt{y}\, \mathtt{R(x, y)} = 1) \vee (\forall \mathtt{z}\, \mathtt{R(a, z)} = 0)$ where $\mathtt{x} \in \mathtt{i}_{\mathtt{obj1}} = \{\mathtt{a, b, c}\}$ and $\mathtt{y, z} \in \mathtt{i}_{\mathtt{obj2}} = \{\mathtt{m, n}\}$ | $\forall_{\mathtt{x}}\exists_{\mathtt{y}}\mathtt{R}_{\mathtt{xy}} + \forall_{\mathtt{z}}\mathtt{R}'_{\mathtt{az}}$ |

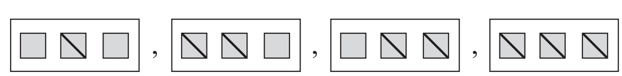
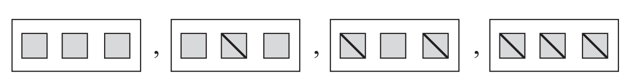
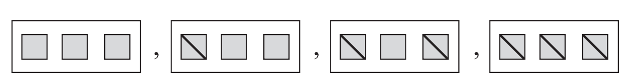
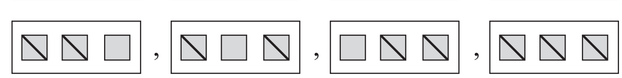
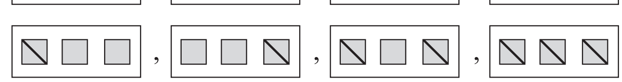
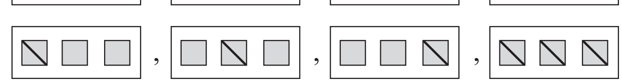
*Note.* The full concept descriptions specify a concept in complete detail. For example, the first description indicates that an item i that contains three objects a, b, and c is an instance of concept $C(\cdot)$ if a and b take value 0 on features F and G, respectively, or if a and c take value 1 on features F and H, respectively. The final description captures a case where each item i specifies a relation R defined over two sets of objects. The summary descriptions in the final column show compact representations of each concept.

ing than simply representing the fact that John reads. Most analyses in this article, however, consider domains where all literals have the same number of places, which means that future studies are needed to explore the assumption that two-place literals are more complex than one-place literals.

The description-length model that relies on language OQ is referred to as the OQ model. Similarly, I refer to the FQ model, the OQ + FQ model, and the propositional model, which is a description length model that relies on the propositional subset of language OQ. The OQ and the propositional models make identical predictions for domains that do not support quantification, but make different predictions when quantification is possible. Comparing these predictions will help to establish whether quantification plays an important role in human concept learning. The OQ, FQ, and OQ + FQ models also lead to different patterns of predictions, and I predict that the OQ model will provide the best account of human learning. Although it may be possible in principle for human learners to quantify over features, the OQ model captures the idea that it is psychologically more natural to quantify over objects than features.

The prediction that the OQ model will perform better than the FQ and OQ + FQ models is motivated in part by previous work which explores the units that humans are inclined to count. The consistent finding is that humans find it natural to count discrete spatio-temporal units, and Spelke-objects (i.e., bounded, coherent physical objects; Spelke, 1990) are the canonical example of these units (Huntley-Fenner, Carey, & Solimando, 2002; Shipley & Shepperson, 1990). For example, if preschoolers are given some red ducks and some green ducks and then are asked to count the number of colors, they often report the number of ducks rather than the number of different colors (Shipley & Shepperson, 1990). Empirical studies also suggest that young children find it easier to count Spelke-objects than parts of these objects (Giralt & Bloom, 2000; Wagner & Carey, 2003). There is no clear consensus in the concept learning literature about what qualifies as an object and what qualifies as a feature, but the entities referred to as objects (e.g., animals, artifacts, and geometric figures) tend to be closer to discrete spatio-temporal units than the entities referred to as features (e.g., colors, textures, and parts such as legs). As a result, the preference for counting discrete spatio-temporal units predicts that

Table 4
*Rules in Language OQ for Several Concepts in Domain (3O,1AF)*

| Type | Rule | *C* | Positive examples |
|---|---|---|---|
| 1 | $\mathtt{F}_{\mathtt{b}}$ | 1 |  |
| 2 | $\mathtt{F}_{\mathtt{a}}\mathtt{F}_{\mathtt{c}} + \mathtt{F}'_{\mathtt{a}}\mathtt{F}'_{\mathtt{c}}$ | 4 |  |
| 4 | $\mathtt{F}'_{\mathtt{b}}\mathtt{F}'_{\mathtt{c}} + \mathtt{F}_{\mathtt{a}}\mathtt{F}_{\mathtt{c}}$ | 4 |  |
| 5 | $\forall_{\mathtt{x}}\exists_{\mathtt{y}}\mathtt{F}_{\mathtt{y}}$ | 1 |  |
| 6 | $(\forall_{\mathtt{x}}\mathtt{F}_{\mathtt{x}}) + \mathtt{F}'_{\mathtt{b}}(\exists_{\mathtt{y}}\mathtt{F}_{\mathtt{y}})$ | 3 |  |
| 10 | $(\forall_{\mathtt{x}}\mathtt{F}_{\mathtt{x}}) + (\exists_{\mathtt{y}}\forall_{\mathtt{z}}\mathtt{F}_{\mathtt{y}}\mathtt{F}'_{\mathtt{z}})$ | 3 |  |

*Note.* The concept types correspond to labels in Figure 2d. The extension of each concept includes four items, and the OQ-complexity of each concept is shown in the column labeled C.

learners will be more inclined to count or quantify over objects than features.

## Other Computational Models of Concept Learning

Psychologists have developed many theories of concept learning, and this section discusses how the rule-based models evaluated in this article relate to these previous contributions. All of the models considered in this section can account for Boolean concept learning to some extent, and a key question is whether they can scale up and account for learning across all of the domains in the conceptual universe.

### Similarity-Based Approaches

Exemplar models propose that learners acquire a concept by storing specific exemplars, and that classification decisions about subsequent exemplars are based on the similarity of these exemplars to the stored exemplars. The similarity between two exemplars is typically formulated as some function of the similarity of these exemplars along individual dimensions such as color and size. The weights assigned to these dimensions need not be fixed, and a similarity-based approach can learn to pay attention to the dimensions that are most informative about category membership. The ALCOVE model (Kruschke, 1992) is one example of this approach, and has been used to account for the relative complexities of concepts in domains 1 and 5 in Table 1 (Kruschke, 1992; Lee & Navarro, 2002; Nosofsky, Gluck, et al., 1994).

The simplest similarity-based models focus on similarity relationships between individual exemplars, but similarity can also be used to organize exemplars into clusters, and classification decisions about subsequent exemplars can be based on their similarities to these clusters. The SUSTAIN model (Love et al., 2004) takes this approach, and has been used to account for the relative complexities of the six concept types studied by Shepard et al. (1961).

Feldman (2000) conducted an extensive study that explores the relative complexities of 76 Boolean concepts, and ALCOVE and SUSTAIN have both been applied to his data. Both models account for Feldman's results to some extent, but neither performs as well as the rule-based models of Feldman (2006) and Goodwin (2006) described in the next section. Developing similarity-based models that account for Boolean concept learning as well as the best rule-based models is therefore an open challenge.

An even greater challenge for the similarity-based approach is to explain how concept learning operates across all of the domains in the conceptual universe. The experiments presented in later sections demonstrate that the complexity of a concept depends in part on whether it can be concisely described using quantification and relations. ALCOVE and SUSTAIN seem unable to account for this result because both models rely on feature-based representations. Other researchers have developed connectionist models that can capture relations (Doumas, Hummel, & Sandhofer, 2008; Hummel & Holyoak, 2003) and can learn to count (Rodriguez, Wiles, & Elman, 1999), and similar techniques may make it possible to develop next-generation versions of ALCOVE and SUSTAIN that account for learning across a large proportion of the conceptual universe. In their current forms, however, these models cannot be applied to all of the domains considered in this article.

### Rule-Based Approaches

Rule-based approaches propose that concepts are represented as rules constructed in a compositional representation language. A key motivation for these approaches is that compositional languages can be used to formulate concepts across all of the domains in the conceptual universe. Different rule-based approaches, however, make different proposals about the nature of the underlying representation language.

Although several authors have pointed out that rules can incorporate quantification and relations (Goodman et al., 2008; Goodwin & Johnson-Laird, 2011; Piantadosi, Goodman, & Tenenbaum, 2010), most of the rule-based models in the literature focus on rules formulated in propositional logic. The rule-plus-exception (RULEX) model proposes that humans learn concepts by constructing conjunctive rules and remembering exceptions to these rules (Nosofsky, Palmeri, & McKinley, 1994). Because any rule in disjunctive normal form corresponds to a collection of conjunctions, the rules considered by RULEX are broadly compatible with the rules considered in this article. In most sections of the article, the predictions of the OQ model do not allow for exceptions, but extending the model in this direction is considered toward the end of the article.

Like the OQ model, the approach of Goodman et al. (2008) proposes that concepts are represented as rules in disjunctive normal form. The representation language considered by these authors is therefore equivalent to the propositional subset of language OQ. Goodman et al. described a probabilistic approach that relies on a prior distribution over rules, and the prior is set up to ensure that shorter rules have higher prior probability. Their approach can therefore be viewed as a probabilistic version of the description length hypothesis.

Goodwin and Johnson-Laird (2011) also considered representations in disjunctive normal form. They referred to each disjunct as a mental model, and proposed that the complexity of a concept corresponds to the number of mental models (or disjuncts) that are required to represent it. Their approach therefore differs from the propositional model evaluated here, which proposes that the complexity of a rule corresponds to the number of literals rather than the number of disjuncts. Although these approaches are distinct, the two are closely related and are motivated by the same basic idea that human learners construct minimal representations in disjunctive normal form.

Feldman (2006) worked with rules that are collections of implications, where each implication takes the form

$$G \leftarrow F_1 \wedge F_2 \wedge \ldots \wedge F_n$$

This implication states that any item with features $F_1$–$F_n$ also has feature G. The *algebraic complexity* of a concept is a function of the minimal set of implications that can be used to characterize it. The default function proposed by Feldman (2006) is equivalent to the number of literals contained in the set of implications, and is therefore an instance of the description length approach. Although the implications considered by Feldman are different from rules in disjunctive normal form, there is an important relationship between these representations that is discussed toward the end of the article.

The four rule-based approaches just described have been applied to a variety of data sets. The analyses most relevant to this article are presented by Feldman (2006) and Goodwin and Johnson-Laird (2011), who both focus on Feldman's data set of 76 Boolean concepts. Feldman's algebraic complexity approach and the mental models approach both account well for the data, but the mental models approach achieves slightly higher quantitative fits. Both models perform better than ALCOVE and SUSTAIN. This article also considers Feldman's data set and shows that the OQ model achieves fits that are roughly comparable to the mental models approach.

Although it is important to show that the OQ model can account for previous studies that focus on Boolean concepts, the more pressing goal here is to explore how people learn concepts that cannot be concisely described in propositional logic. Most previous rule-based models focus on propositional logic, but it should be possible to extend these models in much the same way as the OQ model extends the propositional model evaluated in this article. Any successes achieved by the OQ model should therefore be interpreted more broadly as successes achieved by the rule-based approach to concept learning. The rule-based models considered in this section make different commitments of various kinds, but the similarities between these models far outweigh the differences.

### Representation Versus Process

The discussion of alternative approaches in previous sections focused primarily on the different representations used by these approaches. The OQ model relies critically on predicate logic as a representation language, and stands in sharp contrast to alternatives that do not rely on a compositional representation language. In addition to characterizing the nature of mental representations, a complete cognitive model must also specify the processes that operate over these representations. The OQ model, however, makes minimal claims about cognitive processing. The one assumption required is that cognitive processes are sensitive to the length of representations formulated in language OQ, but there are many ways in which this assumption might be satisfied. This section describes how the OQ model is compatible in principle with the processing assumptions made by several previous accounts of learning and reasoning.

Many previous accounts that rely on predicate logic treat logic as an account of reasoning or inference. These accounts, for example, propose that logical inference helps to explain how humans decide which conclusions follow from a given set of statements (Rips, 1994). The mental models approach was developed in opposition to this view, and proposes that reasoning is better described as a process of constructing and inspecting mental models. The OQ model makes no commitment to either of these views. The model treats predicate logic as an account of mental representation instead of an account of reasoning, and therefore makes no claim about whether human reasoning is better characterized as mental theorem-proving or model-based inference. It seems possible that humans rely on both kinds of inference strategies, and the OQ model is fully compatible with this possibility.

Logical inference is clearly one way in which logical representations could be used, but probabilistic inference is another possibility. Several researchers have discussed how probabilistic inference can operate over logical representations (Goodman et al., 2008; Kemp, Goodman, & Tenenbaum, 2008b; Piantadosi et al., 2010), and the approaches described by these researchers can be used to develop probabilistic models of learning and reasoning that rely on representations formulated in language OQ. Logic and probability are sometimes viewed as competing approaches, and indeed the two may conflict when both are treated as accounts of how people infer what follows from a given set of premises (Oaksford & Chater, 2002). Logic and probability, however, can be naturally combined if predicate logic is treated as an account of knowledge representation and probability theory is used to explain how logical representations are learned and used for inductive inference.

This section has argued that the OQ model is compatible with several different kinds of processing assumptions. The present article focuses on knowledge representation rather than reasoning, and exploring an approach that makes minimal assumptions about processing is the simplest initial way to evaluate the representational merits of language OQ. Ultimately, however, the OQ model will need to be supplemented with a detailed account of processing, and the general discussion outlines some steps that can be taken in this direction.

## Empirical Studies of Concept Learning

Previous researchers have collectively explored concept learning across many domains, but any given study typically considers a single domain or a small handful of domains. A comprehensive account of concept learning should apply across multiple domains, and here I evaluate the OQ model using data from 11 different domains. The set of domains is large by the standards of previous research, but even so it covers only a small part of the conceptual universe. To compensate in part for this limitation the domains are deliberately chosen to include a variety of concepts, including feature-based concepts, relational concepts, and concepts that rely on quantification. Where possible, I analyze previously-published data sets and compare predictions of the OQ model with the best published results for a given problem. Some of the domains, however, have not previously been studied, and I therefore report results for two new experiments.

As described already, the conceptual universe includes real-world categories such as "aromatic hydrocarbons" and "armed robbery" in addition to artificial examples such as categories of geometric stimuli. The analyses in the following sections focus primarily on simple artificial stimuli, because stimuli of this kind provide the simplest initial way to explore how humans learn about systems of objects, features, and relations. The final analysis, however, focuses on kinship systems, and therefore illustrates how the theoretical approach applies to an important real-world domain.

Because the studies that follow consider a relatively large number of domains, the analysis of any single domain is necessarily limited. Only one study is discussed for each domain, and each study uses either average learning times or average learning accuracies to assess the relative subjective complexities of the concepts that can be formulated within a given domain. The data emerging from these experiments are sufficient to distinguish the OQ model from several alternatives, but at least two important aspects of concept learning are left unaddressed. First, different individuals may learn in different ways, and a concept that is difficult for one person may be relatively easy for another to learn. Second, different measures of conceptual complexity sometimes produce different results, and these differences can be informative about the context-dependence of human learning. Although the current analyses do not address either of these factors, the results support the conclusion that predicate logic is a valuable representational substrate for models of human concept learning. It may therefore be useful to consider extensions of the OQ model that account for individual differences and context-sensitive reasoning, and the general discussion outlines some possible steps in this direction.

## Feature-Based Domains Related to Shepard et al. (1961)

The work of Shepard et al. (1961) is undoubtedly the most influential study of conceptual complexity. These researchers characterized the six SHJ types shown in Figure 2b and studied how these types are learned in domains including (1O,3SF) and (3O,3SF) from Table 2. They explored several different ways of measuring conceptual complexity, and found that the same complexity ordering emerged for all measures across all of the domains that they considered. From least to most complex the ordering is I < II < III, IV, V < VI, where types III, IV, and V are of roughly equal complexity.

The family of domains in Table 2 includes several of those that Shepard et al. (1961) considered. As discussed previously, it is natural to expect that the six SHJ types apply across all of these domains, but in reality some of these domains include nine or 10 types. This section describes an experiment that explores four of the domains in Table 2 and tests the prediction that human learners are sensitive to differences that are not captured by the SHJ types.

The four domains considered include (3O,3SF), a domain originally studied by Shepard et al. (1961), along with domains (3O,1AF), (1O,3AF), and (3O,3AF). One goal of the experiment is to provide a comprehensive investigation of the size four concepts that can be formulated within these domains. Figure 2d suggests that the three domains with additive features include 10 distinct types of size 4, and the experiment explores the complete set of 10 types across all of the four domains. Previous studies of domain (3O,1AF) have been organized around the six SHJ types (Goodwin, 2006; Mathy, 2010), but these types do not accurately reflect the different kinds of concepts that can be formulated within this domain. Including the full set of 10 types is necessary in order to draw general conclusions about how people think about the domain.

A second goal of the experiment is to focus specifically on the role that quantification plays in concept learning. The most striking difference between domain (3O,3SF) and the three domains with additive features is that additive features open up the possibility of representations that rely on quantification. For example, type 5 in domain (3O,1AF) includes items where two or more objects have a slash (see Table 4), and type 10 includes items where either one or three objects have slashes. Quantification or counting is possible in principle for all three of the domains with additive features, but the entities available for counting differ across these domains. For example, the items in domain (3O,1AF) include three objects and one feature, and therefore support quantification over objects but not over features. The items in domain (1O,3AF) include one object and three features, and therefore support quantification over features but not over objects. Studying how concept types 5 and 10 are learned within these domains can therefore help to establish whether human learners find it natural to quantify over objects and features. The OQ model supports quantification over objects but not features, and the experiment explores whether this aspect of the model is consistent with human learning.

**Method.**

***Participants.*** Eighty Carnegie Mellon University (CMU) undergraduates participated for course credit. All were naive with respect to the purpose of the experiment.

***Materials.*** The experiment was carried out using a custom built graphical interface. The items in each domain were cards that each included a vertical array of shapes (i.e., objects). In domain (3O,3SF), the first shape was either a purple triangle or a purple circle, the second shape was either an orange square or a green square, and the final shape was either a purple square with a black horizontal band or a purple square with a black vertical band. The items for domains (3O,1AF), (3O,3AF), and (1O,3AF) were similar to the items shown in Table 2 except that the objects on each card were arranged vertically rather than horizontally. In domain (3O,1AF), the three shapes were purple squares where each square either did or did not have a black slash. In domain (3O,3AF), all three shapes were purple squares, and the first, second, and third squares either had or did not have a thick black boundary, a grid of dots, and a slash, respectively. In domain (1O,3AF), each item showed a purple square that either had or did not have a thick black boundary, a grid of dots, and a slash. The three additive features (i.e., boundary, grid, and slash) were based on stimuli developed by Sakamoto and Love (2004).

The 10 concept types in Figure 2d were used to create 10 concepts from each domain. The concepts chosen for each domain were directly comparable. For example, concept 1 in each domain can be described as follows: all cards where the lowest square had a black horizontal band (domain (3O,3SF)); all cards where the lowest square had a slash (domain (3O,1AF)); all cards where the lowest square had a slash (domain (3O,3AF)); all cards where the single square had a slash (domain (1O,3AF)).

***Procedure.*** Each participant was assigned to a single domain and learned all 10 concepts from that domain. Twenty participants were assigned to each domain. For each learning problem in each domain, all eight items were simultaneously presented on screen, and participants were able to drag them around and organize them however they liked. Each problem had three phases: a learning phase, a memory phase, and a description phase. During the learning phase, the four items belonging to the current concept had red boundaries, and the remaining four items had blue boundaries. Participants could press a key at any time to move to the memory phase, where the items were displayed in random positions on the screen, the colored boundaries around the items were removed, and participants were asked to sort the items into the red group and the blue group. If they made an error they returned to the learning phase, and could retake the test whenever they were ready. If they successfully sorted the cards, they proceeded to the description phase and were asked to provide a written description of the two groups of cards.

A timer was visible on screen during the learning phase, and counted down to zero from a starting point of 300 s (5 min). The timer stopped when participants moved to the memory phase, but started again if they made an error and returned to the learning phase. Once the timer had reached zero, participants were allowed to proceed past the memory phase even if they could not successfully sort the items. The maximum time spent learning any concept is therefore 300 s.

Before starting the first of the 10 problems, the procedure was introduced using an introductory problem where the eight items were words drawn from two categories: fruits (banana, peach, apple, pear) and vehicles (truck, car, boat, and plane). All participants completed this introductory problem first, and the order of the remaining 10 concepts was pseudo-randomized across participants using a Latin square. The color assignments (red or blue) were also pseudo-randomized—in other words, the "red groups" learned by some participants were identical to the "blue groups" learned by others.

***Model predictions.*** Table 5 shows predictions about the relative complexities of the concepts in all four domains. The predicted complexity of each concept is based on the length of its minimal description in language OQ and was computed using a method described in Appendix B. Some concepts admit multiple descriptions of minimal complexity, and Table 5 includes one minimal description for each concept.

The minimal descriptions and complexities in Table 5 are based on the concepts that include the gray nodes in Figure 2d. The complexity values in parentheses show the complexities of the complements of these concepts—in other words, concepts that include the white nodes in Figure 2d. In most cases the concepts and their complements have equal complexities, but language OQ predicts that concepts 8 and 9 are simpler than their complements in domains (3O,1AF) and (3O,3AF). Because the experiment used arbitrary labels (i.e., "red items" and "blue items") for positive and negative examples, the complexities for concepts and their complements are averaged to generate the final predictions.

The predicted complexity for each concept in domain (3O,3SF) is identical to its complexity in propositional logic. Because domain (3O,3SF) does not support quantification, language OQ reduces to propositional logic in this case, and the minimal descriptions for domain (3O,3SF) are all propositional rules. Domain (3O,1AF), however, includes several concepts that can be concisely represented using quantifiers. The minimal descriptions for concepts 1, 2, 4, 5, 6, and 10 are identical to the rules in Table 4, and were explained in a previous section.

Domains (3O,1AF) and (3O,3AF) present an interesting contrast. Because the same feature *F* applies to all three objects in domain (3O,1AF), rules such as $\forall_x F_x$ can be formulated. In domain (3O,3AF), however, each feature applies to only one object, which means that rules such as $\forall_x F_x$ are not possible. In domain (3O,3AF), however, each item has a characteristic pattern (slash for object a, spots for object b, and frame for object c), which opens up the possibility of statements such as "all objects have their characteristic patterns." Language OQ can capture statements of this kind if we introduce an indicator feature I that indicates whether each object has its characteristic pattern. The model predictions in Table 5 make use of this indicator feature, and the resulting complexity predictions are identical to those for domain (3O,1AF). For example, the shortest description for concept 5

Table 5
*Minimal Descriptions and OQ-Complexities for the 10 Concepts in Experiment 1*

| Type | (3O, 3SF) | *C* | (3O, 1AF) | *C* |
|---|---|---|---|---|
| 1 | $G_b$ | 1 (1) | $F_b$ | 1 (1) |
| 2 | $F'_aH'_c + F_aH_c$ | 4 (4) | $F'_aF'_c + F_aF_c$ | 4 (4) |
| 3 | $F'_aG_b + F_aH_c$ | 4 (4) | $F'_aF_b + F_aF_c$ | 4 (4) |
| 4 | $G'_bH'_c + F_aH_c$ | 4 (4) | $F'_bF'_c + F_aF_c$ | 4 (4) |
| 5 | $F_aG_b + G_bH_c + F_aH_c$ | 6 (6) | $\forall_x\exists_y F_y$ | 1 (1) |
| 6 | $F_aG'_b + G'_bH_c + F_aH_c$ | 6 (6) | $(\forall_x F_x) + F'_b(\exists_y F_y)$ | 3 (3) |
| 7 | $F_aG'_bH'_c + F'_aH_c + G_bH_c$ | 7 (7) | $(\forall_x F_x) + F'_aF_c + F_a(\exists_x\forall_y F'_y)$ | 5 (5) |
| 8 | $F_aG_bH_c + F'_aG'_b + G'_bH'_c$ | 7 (7) | $(\forall_x F_x) + F'_b(\exists_x\forall_y F'_y)$ | 3 (4) |
| 9 | $F'_aG'_bH'_c + F_aG_b + G_bH_c$ | 7 (7) | $(\forall_x F'_x) + F_b(\exists_x\forall_y F_y)$ | 3 (4) |
| 10 | $F_aG'_bH'_c + F'_aG_bH'_c + F'_aG'_bH_c + F_aG_bH_c$ | 12 (12) | $(\forall_x F_x) + (\exists_y\forall_z F_yF'_z)$ | 3 (3) |
| Type | (1O, 3AF) | *C* | (3O, 3AF) | *C* |
| 1 | $G_a$ | 1 (1) | $I_b$ | 1 (1) |
| 2 | $F'_aH'_a + F_aH_a$ | 4 (4) | $I'_aI'_c + I_aI_c$ | 4 (4) |
| 3 | $F'_aG_a + F_aH_a$ | 4 (4) | $I'_aI_b + I_aI_c$ | 4 (4) |
| 4 | $G'_aH'_a + F_aH_a$ | 4 (4) | $I'_bI'_c + I_aI_c$ | 4 (4) |
| 5 | $F_aG_a + G_aH_a + F_aH_a$ | 6 (6) | $\forall_x\exists_y I_y$ | 1 (1) |
| 6 | $F_aG'_a + G'_aH_a + F_aH_a$ | 6 (6) | $(\forall_x I_x) + I'_b(\exists_y I_y)$ | 3 (3) |
| 7 | $F_aG'_aH'_a + F'_aH_a + G_aH_a$ | 7 (7) | $(\forall_x I_x) + I'_aI_c + I_a(\exists_x\forall_y I'_y)$ | 5 (5) |
| 8 | $F_aG_aH_a + F'_aG'_a + G'_aH'_a$ | 7 (7) | $(\forall_x I_x) + I'_b(\exists_x\forall_y I'_y)$ | 3 (4) |
| 9 | $F'_aG'_aH'_a + F_aG_a + G_aH_a$ | 7 (7) | $(\forall_x I'_x) + I_b(\exists_x\forall_y I_y)$ | 3 (4) |
| 10 | $F_aG'_aH'_a + F'_aG_aH'_a + F'_aG'_aH_a + F_aG_aH_a$ | 12 (12) | $(\forall_x I_x) + (\exists_y\forall_z I_yI'_z)$ | 3 (3) |

*Note.* The complexity values in parentheses show the complexities of the complements of the 10 concepts in Figure 2d. For domain (3O,3AF), the indicator feature I indicates whether each object has its characteristic pattern. The rows are grouped into blocks that correspond to the six SHJ types—for example, concepts 3 and 4 both belong to SHJ type III.

indicates that all members of the concept have two or more objects with patterns.

The indicator feature I in Table 5 goes beyond the specification of domain (3O,3AF) in Table 2, and assumes that learners are able to align the three features shown there ($F_1$, $F_2$, and $F_3$). Alignment is possible in this case because each feature is additive. Feature alignment may also be possible in domains that do not rely on additive features. For example, learners may find it natural to align a size feature (big/small) with a volume feature (loud/soft) because the values of both features indicate quantities of some kind (L. Kotovsky & Gentner, 1996). Characterizing the full set of cases where features can be aligned is a challenge for future work, and addressing this challenge will help to explain how people think about the full set of domains in the conceptual universe (see Table 2). For present purposes, I simply adopt the working hypothesis that additive features can be aligned.

Domain (1O,3AF) in Table 5 includes a single object only, and quantification over objects captures no useful information in this case. The predicted complexities for domain (1O,3AF) are therefore identical to the propositional complexities predicted for domain (3O,3SF).

Several qualitative predictions can be extracted from the quantitative predictions in Table 5. First, the OQ model predicts that domains (3O,1AF) and (3O,3AF) include concepts that belong to the same SHJ class (e.g., concepts 5 and 6) but have different subjective complexities. Second, type VI is traditionally considered to be the most difficult of the six SHJ types, but the OQ model predicts that this type (which corresponds to type 10 in Table 5) is not the most difficult type within domains (3O,1AF) and (3O,3AF). Third, the OQ model predicts that humans find it natural to quantify over objects, and therefore predicts that the two domains which support quantification over objects ((3O,1AF) and (3O,3AF)) will produce different results than the domains which do not support quantification over objects. Finally, the OQ model predicts that humans do not find it natural to quantify over features, and predicts that a domain which supports this kind of quantification (domain (1O,3AF)) will produce the same kind of results as a domain that does not support quantification over features (domain (3O,3SF)).

**Results.** The computer interface recorded the length of time participants spent on the learning phase for each concept. Total learning times for the 10 concepts suggested that domain (3O,1AF) was the easiest domain, and domain (3O,3SF) was the most difficult. On average, participants took 397, 509, 546, and 667 s to learn all 10 concepts in domains (3O,1AF), (3O,3AF), (1O,3AF), and (3O,3SF), respectively (standard deviations are 397, 250, 210, and 149 s).

Because the total learning times varied widely within each domain, all remaining analyses will use learning times that are normalized to sum to 1 for each participant. Normalizing in this way factors out individual differences in speed and provides a sensitive measure of the relative difficulties of the concepts within each domain. One disadvantage of working with normalized learning times is that these normalized times cannot be directly compared across domains. For example, the normalized learning time for concept 5 in domain (3O,3AF) cannot be compared with the normalized learning time for concept 5 in domain (1O,3AF). Comparisons of this kind, however, are unlikely to be informative because the domains are not matched for factors including perceptual complexity. Choosing stimuli that are matched in this way would be ideal but is unlikely to be possible. For example, domains (3O,3AF) and (1O,3AF) are matched in the sense that they use exactly the same features, but in domain (1O,3AF) these features occupy the same spatial location, which presumably increases perceptual load and makes the items in this domain more difficult to process. As a result, it is meaningful to compare the complexity profile over the 10 concepts in domain (3O,3AF) with the corresponding profile for domain (1O,3AF), but less meaningful to compare the complexities of individual concepts within each domain.

Figure 4a shows the mean normalized learning times for each domain, and indicates the relative difficulties of the concepts within each domain. The results in Figure 4a support the qualitative predictions identified previously. Domains (3O,3SF) and (1O,3AF) produce results that are broadly consistent with the standard complexity order over the six SHJ types. Domains (3O,1AF) and (3O,3AF), however, produce results that differ qualitatively from the SHJ type ordering. For example, type 10 (which corresponds to SHJ type VI) is not the most difficult type in domain (3O,1AF) or domain (3O,3AF).

The results also suggest that pairs of concepts which belong to the same SHJ type can differ in subjective complexity. There are five such pairs: concepts 3 and 4, 5 and 6, 7 and 8, 7 and 9, and 8 and 9. The model predicts that differences within these pairs will be small or non-existent in domains (3O,3SF) and (1O,3AF), and paired-sample *t*-tests indicate that none of these differences is statistically significant. Within domain (3O,1AF), the model successfully predicts that concept 5 is easier to learn than concept 6, and that concepts 8 and 9 are easier to learn than concept 7. As shown in Figure 4a, two out of these three differences are statistically significant ($p < .001$ and $p < .01$). In domain (3O,3AF), the learning times for concepts 8 and 7 are not significantly different, but concept 5 is easier to learn than concept 6 ($p < .01$).

Figure 5 compares human learning times with complexity predictions according to four languages: propositional logic, language OQ, language FQ, and language OQ + FQ. Domain (3O,3SF) does not support quantification over objects or features, and all four languages therefore generate equivalent propositional rules as minimal descriptions of the 10 concepts. Domains (3O,1AF) and (3O,3AF) open up the possibility of object quantification, and the languages that support object quantification account best for the results for these domains. Domain (1O,3AF) opens up the possibility of feature quantification, but the two languages which support feature quantification (FQ and OQ + FQ) account poorly for learning times within this domain. Language OQ is the only language that successfully accounts for the result from all four domains, suggesting that humans find it natural to quantify over objects but not features.

Correlations between each set of model predictions and learning times are shown in Figure 5, and the interval below each correlation is a 95% confidence interval generated by bootstrapping at the level of individual participants. All models achieve identical correlations for domain (3O,3SF), and for each of the three remaining domains the correlation achieved by the two best models for that domain is significantly higher than the correlation achieved by the remaining two models. The relevant confidence intervals do not overlap for domains (3O,1AF) and (1O,3AF), indicating that the difference between correlation coefficients is significant at the 0.05 level. The confidence intervals do overlap for domain (3O,3AF), but

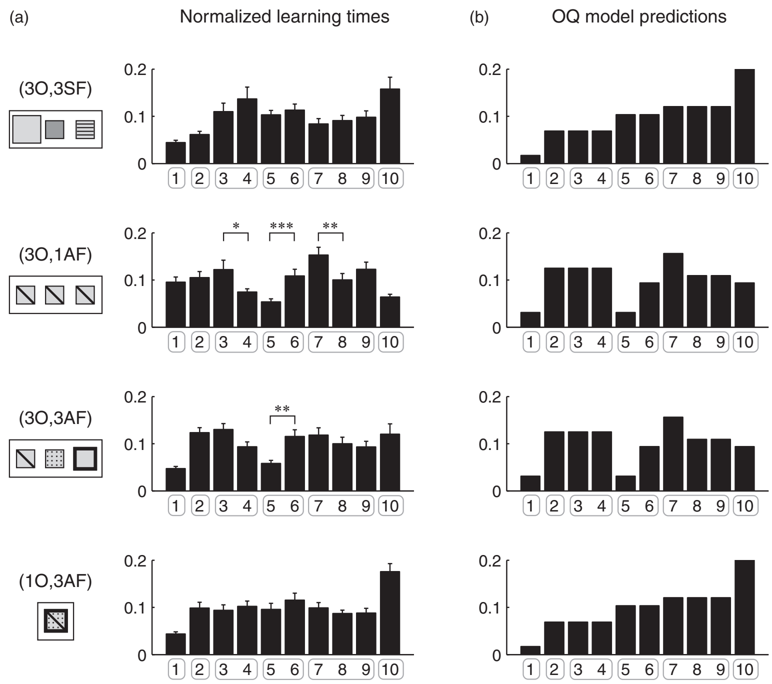



*Figure 4.* (a) Normalized learning times for the 10 concepts in Experiment 1. The six ovals along the *x*-axis group concepts that belong to the same SHJ type—for example, concepts 3 and 4 both belong to SHJ type III. (b) Normalized OQ-complexity values. $^{*}$ $p < .05$. $^{**}$ $p < .01$. $^{***}$ $p < .001$.

the correlation achieved by models OQ and OQ + FQ exceeded the correlation achieved by the other models for more than 95% of the bootstrap samples, indicating that the difference between correlation coefficients is significant at the 0.05 level.

Any rule in OQ is also a rule in language OQ + FQ, and the comparison between the OQ and OQ + FQ models is therefore especially interesting. The two models generate identical minimal descriptions for all domains except (1O,3AF). Domain (1O,3AF) includes several concepts that can be concisely described using quantification over features. For example, concept 5 includes "all items where the square has two or more of the three features." The OQ + FQ model takes advantage of this possibility and predicts that concept 5 is relatively easy to learn in domain (1O,3AF). The OQ model cannot quantify over features, and therefore generates a minimal description for concept 5 that is substantially longer. The results in Figure 4a indicate that concept 5 is not especially easy in domain (1O,3AF), and therefore suggest that language OQ + FQ is too expressive to accurately capture human learning. Language OQ accounts better for the data, suggesting that the constraints captured by this language may correspond to constraints that shape how mental representations are constructed.

The written descriptions generated by participants provide additional evidence that many of them relied on quantification over objects. None of the domain (3O,3SF) participants described a criterion for distinguishing between red and blue cards that referred to quantification over objects. In domain (3O,1AF), however, 15 out of 20 descriptions of concept 5 and 13 out of 20 descriptions of concept 10 referred to quantification over objects. One representative description of concept 5 stated that "the red cards have fewer than 2 slashed rectangles" and "the blue cards have at least 2 slashed rectangles." A representative description of concept 10 indicated that "the red group has 1 slash or 3 slashes" and that "the blue group has either 2 slashes or 0 slashes," and another mentioned that "the blue cards have an even number of slashed rectangles." As mentioned earlier, concept 10 corresponds to SHJ type VI. Shepard et al. (1961) previously noted that type VI can be learned by distinguishing between odd counts (1 or 3) and even counts (0 or 2), and the written descriptions suggest that this strategy is especially natural within domain (3O,1AF).

For domain (3O,3AF), 14 out of 20 descriptions of concept 5 and 11 out of 20 descriptions of concept 10 referred to quantification or counting. One representative description of concept 5 stated that "reds have two or more patterned boxes." The corresponding counts for domain (1O,3AF) were substantially smaller: 4 out of 20 descriptions of concept 5 and 3 out of 20 descriptions of concept 10 referred to quantification or counting. One of the few descriptions that referred to quantification indicated that red items had "multiple features," and blue items had "only one feature" (concept 5), and the same participant indicated that "red = even number of features," and "blue = odd number of features" (concept 10). Descriptions like these suggest that people can count or quantify over features, but the relative rarity of these descriptions suggests that it is psychologically more natural to quantify over objects rather than features.

Thus far, I have focused on differences in learning patterns across the four domains in Figure 4 and argued that the predictions of the OQ model are broadly compatible with the main differences that emerge. The model, however, makes some specific predictions within individual domains that are not supported and that highlight some of its limitations. Two discrepancies between model predictions and the

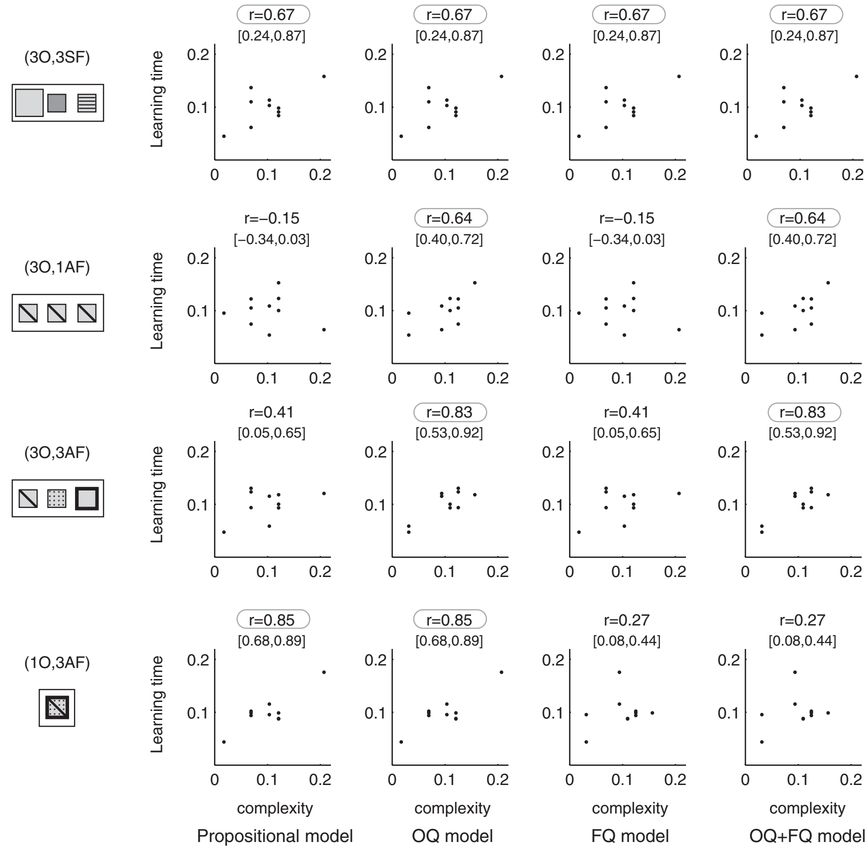



*Figure 5.* Normalized learning times for Experiment 1 plotted against normalized complexity values predicted by four languages: propositional logic, OQ, FQ, and OQ + FQ. Many of the plots are identical because some pairs of languages are effectively equivalent within some domains. For example, quantification is not possible within domain (3O,3SF), which means that the final three languages reduce to propositional logic within this domain. The highest correlations for each domain are circled, and the interval below each correlation is a 95% confidence interval.

human data are especially apparent. First, concept 2 in domain (3O,3SF) is easier than concepts 3 and 4, although the OQ model predicts no difference in learning times across these concepts. The empirical result here is consistent with previous studies which have documented that SHJ type II is subjectively less complex than SHJ type III (Shepard et al., 1961). A classic explanation of this result is that type II can be described by referring to two features (F and H in Table 5), but the simplest description for type III refers to all three features (Shepard et al., 1961). A preference for rules that refer to fewer features is consistent with the key psychological notion of selective attention (Kruschke, 1992), but is not captured by the complexity measure used in this article. Some previous rule-based models are limited in the same way—note that Boolean complexity and the mental models account both predict that types II and III are equally difficult (Goodwin & Johnson-Laird, 2011; Lafond, Lacouture, & Mineau, 2007; Vigo, 2006). Nosofsky, Palmeri, and McKinley (1994) and Goodman et al. (2008), however, have presented rule-based approaches that can account for the difference in complexity between types II and III. A more sophisticated version of the description-length approach should also be able to account for this result. For simplicity, the OQ model assumes that the complexity of a rule is determined only by the number of symbols that it contains, but a more principled description-length account would take the identity of these symbols into account and would use shorter codes for frequently encountered symbols (Grünwald, 2007). As a result, a rule with repeated symbols would have a shorter description than a rule where all symbols are different even if the two rules are matched for overall length.

A second limitation of the model is that it fails to predict that concept 3 in domain (3O,1AF) is more difficult than concept 4. Because concepts 3 and 4 are both instances of SHJ type III, most existing models of concept learning would also fail to predict this difference. Many of the written descriptions of concept 4 mentioned that the items in one group can be organized into a progression that starts with the item with no slashes and finishes with the item with three slashes, and where one slash is added with each successive item. The positive instances of concept 4 in Table 4 are arranged from left to right according to this progression. This progression can be viewed as a global property of concept 4 that emerges only when all positive instances of the concept are considered simultaneously. The model is not sensitive to global properties of this kind, and focuses instead on learning a criterion of concept membership that can be applied to any given item in isolation. Although humans are sometimes sensitive to global properties such as the sequential progression just described, whether or not these properties are noticed is likely to depend on the experimental paradigm used. It is possible that concepts 3 and 4 would be equally difficult in the context of a learning paradigm where only one item is visible at any time.

**Discussion.** Although the model fails to account for some aspects of the data, the overall pattern of results supports three general conclusions. First, there is a clear qualitative difference between the results for domains (3O,3SF) and (1O,3AF) and the results for domains (3O,1AF) and (3O,3AF). Even though all four domains may initially appear to be isomorphic, the behavioral results support the analysis in Figure 2 which identifies different conceptual types within these domains. Second, the differences in learning times across the four domains are broadly consistent with the hypothesis that the mental representations of some concepts rely on quantification or counting. In particular, concepts 5 and 10 within domain (3O,1AF) can be concisely expressed using quantification, and the description length model successfully predicts that these concepts are relatively easy to learn. Third, the results suggest that it is psychologically more natural to quantify over objects than features. All three conclusions are consistent with language OQ, which accounts relatively well for learning times across all four domains considered.

Quantification plays an obvious role in natural language and previous studies have explored how humans learn and think about quantifiers such as "for all" and "there exists" (Brooks & Braine, 1996; Johnson-Laird & Byrne, 1991; Szymanik & Zajenkowski, 2010). Relative to this previous work, the contribution of Experiment 1 is to explore the role that quantification plays in concept learning. Previous accounts of conceptual complexity typically rely on representations that are similar to some form of propositional logic, but Experiment 1 suggests that any comprehensive account of conceptual complexity should also take quantification into account.

Although the written descriptions support the general prediction that participants rely on quantification over objects, they also expose a limitation of the model. Language OQ uses universal and existential quantifiers to describe concepts 5 and 10, but participants often describe these concepts using numbers. Both kinds of descriptions are similar in one important respect, because both rely on specifications of quantity. Ultimately, however, a complete account of concept learning should include some psychologically realistic proposal about how numbers are mentally represented (Carey, 2009). Language OQ was able to account for the data from domain (3O,1AF) because it supports concise descriptions of statements involving small numbers, but this language becomes increasingly unwieldy as the numbers involved increase in size. Consider, for example, the variant of domain (3O,1AF) where each item includes 10 objects. The concept including all items with five or more slashes seems conceptually simple, but the minimal OQ representation is relatively complex: $\exists v \exists w \exists x \exists y \exists z F_v F_w F_x F_y F_z$. This article focuses on language OQ because this language is the natural next step beyond propositional logic, but future studies can explore richer languages that can directly express numerical statements. In particular, richer representation languages will be needed in order to explain how people learn concepts that rely on relatively large numbers.

Because the written descriptions are often informative, it is possible that the mental representations used for learning the 10 concepts are formulated in a natural language such as English. The data suggest that the underlying representation language incorporates quantification over objects, and English is one representation language that satisfies this criterion. Note, however, that a simple description length approach will not account for the data if English is the underlying representation language. Like language OQ + FQ, English supports both object and feature quantification, which means that concepts 5 and 10 have short English descriptions in both domain (3O,1AF) and domain (1O,3AF). If English is the underlying representation language, then some additional principle needs to be invoked to explain why quantification over objects is more natural than quantification over features.

Although Experiment 1 suggests that participants are more likely to rely on quantification over objects than quantification over features, there are good reasons to believe that quantification over features or dimensions is part of the human conceptual repertoire. Some everyday concepts appear to rely on quantification over features or dimensions (Sassoon, 2011). A teacher may classify a child as a "prodigy" if her performance is exceptionally good along some dimension, and as "abnormal" if there is some dimension along which she is far from average. A nurse who is examining a group of children to determine which children have spots, which children have fever, and which children have joint pain may classify any child as "sick" who has one or more of these symptoms. Relational concepts such as "same" or "different" are also naturally formulated using quantification over features or dimensions: for example, two objects are different if there is some feature that distinguishes between the two. Examples like these suggest that it may be valuable to explore why quantification over features is exploited in some contexts but not others. The nature of the features involved is likely to be relevant—for example, it may be more natural to quantify over features like "has glasses" and "has an earring" than features like "has a rash," because features like "has glasses" can be conceptualized in terms of the presence or absence of discrete spatio-temporal units. It is also possible that paradigms that allow participants to compare pairs of objects are more likely to elicit quantification over features than paradigms where the stimulus objects are presented serially. The process of comparing two objects may lead participants to count the respects in which they are different, and the fact that comparisons of this sort were possible in Experiment 1 may help to explain why a small number of participants gave written descriptions that explicitly referred to quantification over features.

The results in Figure 4a are broadly consistent with results from previous studies that have explored domains similar to (3O,1AF). One general theme that emerges is that SHJ type VI, which corresponds to type 10, is not the most difficult type in this domain. Mathy (2010) explored a domain where each item included three black balls and a horizontal line, and where each ball was either above or below the line. Mathy focused on SHJ types IV and VI, and found that type VI was easier than type IV. His study considered multiple instances of type IV that were chosen at random and therefore included instances of types 5 and 6 in Figure 2d. His results are therefore consistent with the data in Figure 4a, which show that the average learning time for concepts 5 and 6 exceeds the average learning time for concept 10.

Goodwin (2006) explored a domain where each item included three switches, and each switch was set either to the left or the right. He considered SHJ types III–VI, and chose instances of these types that correspond to types 3, 5, 8, and 10, respectively. Consistent with the data in Figure 4a, he found that concept 3 was more difficult than the remaining three concepts. Note, however, that type 4 in Figure 4a is an instance of Type III and is relatively easy to learn, which challenges Goodwin's conclusion that type III is uniformly more difficult in the context of domain (3O,1AF).

Although previous studies have explored domains similar to (3O,1AF), they have not acknowledged that this domain contains 10 types and therefore have not documented the relative complexities of these types. The design of Experiment 1 was made possible by the formal characterization of the conceptual universe, and the results suggest that this characterization can provide a useful foundation for empirical studies and computational modeling. For example, the set-theoretic characterization of domain (3O,1AF) led to the distinction between types 5 and 6 in Figure 2d, and the same basic approach can be used to identify the full set of cases that must be considered in any domain.

## Boolean Concepts

Experiment 1 built on the work of Shepard et al. (1961) by exploring concept learning across four qualitatively different domains. Feldman (2000) extended the work of Shepard et al. in a different direction by systematically exploring two closely-related domains that I refer to as (1O,3SF) and (1O,4SF). Table 2 shows one possible instantiation of domain (1O,3SF), and domain (1O,4SF) is similar except that each item in the domain now includes an additional substitutive feature. A recent study by Vigo (2011) includes data for domain (1O,2SF) in addition to the two domains considered by Feldman, but here I analyze Feldman's data in order to compare with models that have previously been evaluated on this data set.

Shepard et al. (1961) focused on concepts of size four, but Feldman systematically explored concepts of different sizes. In domain (1O,3SF), he considered all types of sizes 2, 3, 4, 5, and 6. In domain (1O,4SF) he considered all types of sizes 2, 3, 4, 12, 13, and 14. For each concept considered, positive examples of the concept were presented in the upper half of the screen and labeled as "Examples," and negative examples appeared in the lower half of the screen and were labeled as "NOT examples." Each concept was displayed for $5n$ seconds, where $n$ is the either the number of positive examples or the number of negative examples, depending on which is smaller. After the training period had finished, participants viewed all items in the domain one by one and had to indicate whether each one was a positive or a negative example of the concept. The dependent measure is the proportion of items classified correctly.

Several computational approaches have been applied to Feldman's data set (Feldman, 2000; Vigo, 2009), including the algebraic complexity (Feldman, 2006) and mental models (Goodwin & Johnson-Laird, 2011) approaches previously described. Here, I compare the OQ model to the mental models approach, which has achieved the best published results on Feldman's data set. As mentioned already, the mental models approach relies on a representational scheme that is identical to the propositional subset of language OQ. In the context of Feldman's experiment, the OQ model and the mental models approach are therefore closely related, and the only real difference is that the two rely on different complexity measures. The OQ model measures the complexity of a rule by counting the total number of literals that it contains, but the mental models approach counts the total number of disjuncts in the rule.

Model fits for the mental models approach are shown in the left column of Figure 6. The labels for each plot use notation introduced by Feldman (2000). For example, the plot for concept family 3[2] shows results for all types from the three feature domain (1O,3SF) where either the total number of positive examples or the total number of negative examples is 2. Goodwin and Johnson-Laird (2011) reported model fits for analyses that combine data from all of the plots in Figure 6, but separating these plots is preferable because concepts from different families were presented for different lengths of time during training. Figure 6 shows that the mental models approach achieves relatively good correlations across all of the families. Note that complexity is inversely related to accuracy, which means that perfect performance corresponds to a correlation of $-1$.

If the OQ model is applied in the same way as the mental models approach, the correlations achieved are shown in the middle column of Figure 6. The OQ model performs substantially worse than the mental models approach, but the right column shows that the model performs better if adjusted to capture two strategies that some participants may have used. First, some participants may have chosen to encode the negative examples rather than the positive examples in cases where the positive examples outnumbered the negative examples. Second, some participants may have relied on a brute force strategy to memorize a set of examples instead of identifying the shortest possible representation in language OQ. Appendix C describes how both strategies were formalized in order to generate the results in the right column of Figure 6. The same general approach can be used to adjust the mental models approach, and the resulting correlations are listed in the caption of Figure 6 and are comparable to the correlations shown in the left column of the figure. Overall, then, the mental models approach performs better than the OQ model if neither is adjusted, but the two achieve comparable fits when the OQ model is adjusted to allow for the specific nature of Feldman's experiment.

The same adjustment could be used to generate predictions of the OQ model for all analyses in this article, but in most cases this adjustment does not affect the model predictions. In Experiment 1, for example, all of the concepts in a given family had the same number of positive examples and the same number of negative examples. The strategy of memorizing the smaller set therefore does not apply, and the strategy of brute-force memorization does not affect the relative learnabilities of the concepts because all had the same number of positive examples. For simplicity, all figures except Figure 6 will therefore include unadjusted predictions only, and the correlations produced by adjusting the model predictions are listed in Table C1 in Appendix C.

Although the adjusted OQ model and the mental models approach perform similarly on Feldman's data set, the two approaches make qualitatively different predictions in other contexts. The mental models approach predicts that all Boolean

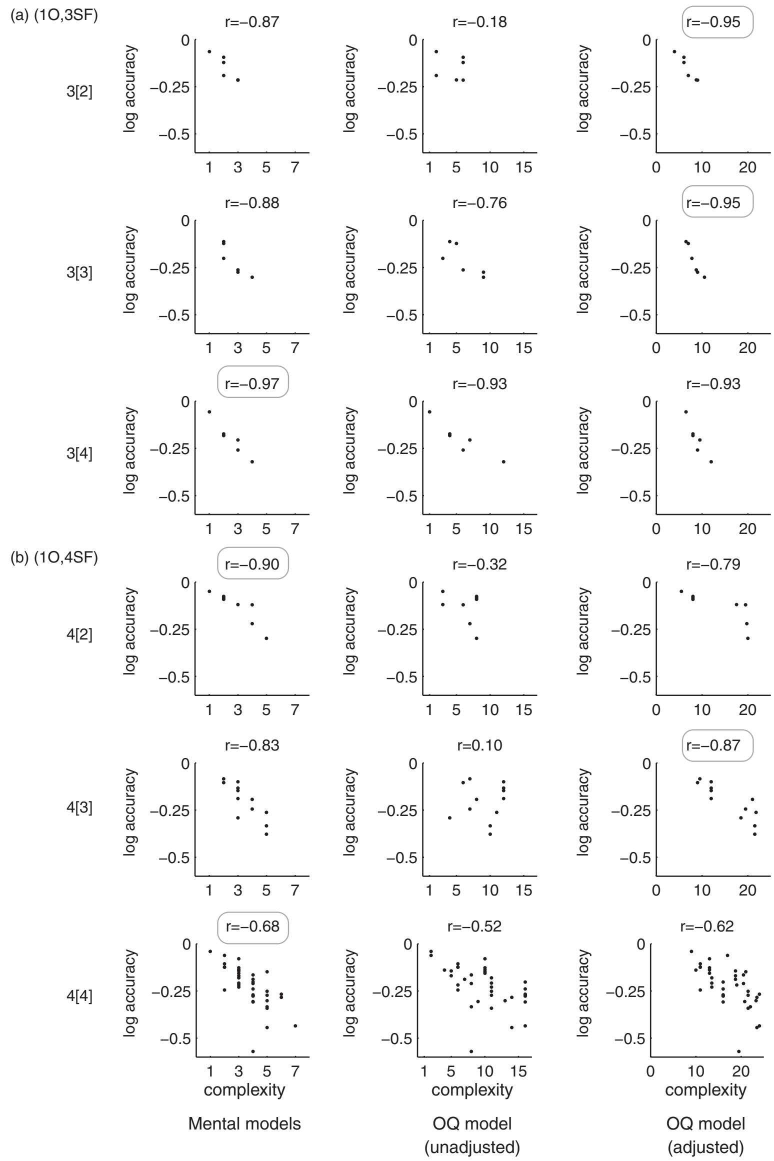



*Figure 6.* Empirical results and model predictions for the six concept families in Feldman's (2000) data set. The three columns show predictions of the mental models approach, the unadjusted OQ model, and the adjusted OQ model. The mental models approach can also be adjusted, and the resulting correlations for the six families of concepts are −0.98, −0.91, −0.97, −0.79, −0.88, and −0.57.

concepts including a single positive example are equally easy to learn regardless of the number of features that the example possesses. For instance, the approach predicts that a concept of size 1 should be equally easy to learn regardless of whether it belongs to family 3[1] or family 10[1]. In contrast, the OQ model predicts that single-item concepts become more difficult to remember as more features are added to the domain. Additional studies are needed to explore differences of this kind and to determine whether minimizing the total number of literals or minimizing the total number of disjuncts provides the better account of human learning across the full space of Boolean concepts.

Here, however, the primary focus is on concepts that go beyond Boolean concepts in various ways, and the key challenge for mental model theory is whether it will be able to account for concept learning across the entire conceptual universe. An important limitation of the theory is that it does not currently explain how learners acquire concepts that rely on quantification and relations, including type 5 in domain (3O,3SF; see Table 4). Mental model theory is often used to explain how people draw inferences from statements involving quantification (Bucciarelli & Johnson-Laird, 1999), but current versions of the theory do not explain how people introduce quantifiers to simplify their mental representations of a situation. Goodwin and Johnson-Laird (2011) acknowledged this limitation but suggested that future versions of the theory may be able to explain how people acquire mental representations that incorporate quantification and relations. At present, then, the OQ model provides broader coverage of the conceptual universe than does mental model theory, but future work on mental model theory may be able to expand its explanatory scope.

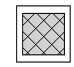
## Ternary Features

Like most previous studies of conceptual complexity, the studies considered so far all focus on binary features, or features that take two values. The conceptual universe, however, includes many domains where features take three or more values. Domain 5 in Table 1 is one example that has been studied by Lee and Navarro (2002) and Aitkin and Feldman (2006). This section discusses the results presented by Aitkin and Feldman and compares them with predictions of the OQ model.

Domain 5 in Table 1 includes nine items in total, and a domain specification is provided in Figure 7a. Figures 7c and 7d show the four concept types of size three and the five concept types of size four. For example, the first concept of size three includes the three items with diagonal crosshatching. It is convenient to represent the three values for each feature as 0, 1, and 2. For example, F(a) = 0 or F(a) = 1 or F(a) = 2, where a is the single object in the domain, and F is the first feature. The third example in Table 3 shows how a concept can be described using these three-valued features. The full concept description includes equality statements like F(a) = 1 and inequality statements like F(a) ≠ 1. For binary features, F(a) ≠ 1 is equivalent to F(a) = 0, which means that inequality statements contribute nothing new to the language. For ternary features, however, inequality statements are qualitatively different from equality statements. Because the domain of interest includes a single object only, F(a) = 1 can be written concisely as $F_1$ and F(a) ≠ 1 can be written as $F'_1$. The summary description in the third row of Table 3 follows this convention.

Table 6 shows the minimal descriptions of the nine concepts in Figures 7c and 7d. The complexity values *C* indicate the OQ-complexities of the concepts, and the values in parentheses indicate the complexities of the complements of the concepts. The complement of the third concept of size 4 has complexity five, and a minimal description for this complement is $F_3 + F_1G_3 + F_2G_2$. All other concepts have the same complexity as their complements.

Table 6
*Minimal Descriptions and Complexity Values for the Domain With Ternary Features*

| Size | Type | (1O, 2SF)  | *C* |
|---|---|---|---|
| (a) Size 3 | 1 | $F_1$ | 1 (1) |
| | 2 | $F_1G'_3 + F_2G_1$ | 4 (4) |
| | 3 | $F_1G'_3 + F_2G_3$ | 4 (4) |
| | 4 | $F_1G_1 + F_2G_3 + F_3G_2$ | 6 (6) |
| (b) Size 4 | 1 | $F'_3G'_3$ | 2 (2) |
| | 2 | $F_1 + F_2G_1$ | 3 (3) |
| | 3 | $F_1G'_3 + F_2G'_2$ | 4 (5) |
| | 4 | $F_1G'_3 + F'_1G_3$ | 4 (4) |
| | 5 | $F_1G'_3 + F_2G_3 + F_3G_1$ | 6 (6) |

*Note.* (a) Minimal descriptions for the three-item types. If a is the single object within the domain, $F_1$ is equivalent to $F_a = 1$, and $F'_1$ is equivalent to $F_a \neq 1$. The values in parentheses show the OQ-complexities of the six-item concepts that are complements of the three-item concepts in Figure 7c. (b) Minimal descriptions and complexity values for the four-item types.

Figure 8 shows the empirical complexity values for each concept reported by Aitkin and Feldman (2006). The task used was similar to the task that generated Feldman's (2000) data set. During training, all items in the domain were presented on screen. Positive examples of the current concept were shown in the upper half of the screen and labeled "In the category." Negative examples were located in the lower half of the screen and labeled "Not in the category." After a fixed presentation time, the nine items were presented one by one and participants indicated whether or not each one was a positive example. The results in Figure 8 are reproduced from Aitkin and Feldman's study and show the mean

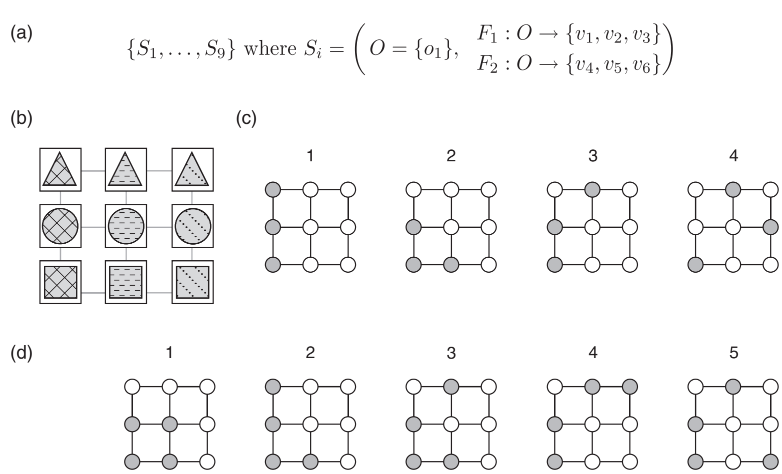



*Figure 7.* (a) A specification of a domain (1O,2SF) that includes two ternary features. (b) A grid showing the nine items in the domain. (c) The four types of size three within the domain. (d) The five types of size four.

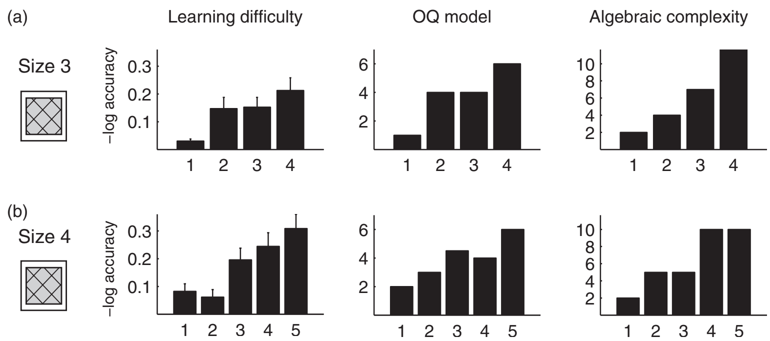



*Figure 8.* Empirical results and model predictions for the domain with ternary features. The plots in the first column are based on results from Figures 3 and 4 in Aitkin and Feldman (2006) and show how difficult the concept types in Figures 7c and 7d were for participants to learn. The remaining columns show predictions according to the OQ model and Feldman's algebraic complexity model.

log proportion correct. The types in Figures 7c and 7d are of sizes three and four, respectively, but some participants learned concepts of sizes six and five that were complements of these concepts. The results reported by Aitkin and Feldman, and replotted in Figure 8, are collapsed over these presentations, and the predictions of the OQ model are therefore based on averages of the complexities of the concepts and their complements.

The model accounts for the broad trends in the data and successfully predicts the easiest and the most difficult concept in each set. Figure 8 shows that the model results are comparable to the algebraic complexity results reported by Aitkin and Feldman (2006). Unlike algebraic complexity, the OQ model successfully predicts that concepts 2 and 3 in the size 3 set are of roughly equal complexity, and that concept 2 is less complex than concept 3 in the size 4 set. Unlike the OQ model, algebraic complexity successfully predicts that concept 3 is less complex than concept 4 in the size 4 set. Following Aitkin and Feldman, the algebraic complexity results are based on the assumption that participants encoded either the positive or the negative examples, depending on which set was smallest. If the same approach is used to adjust the OQ model results, concepts 3 and 4 in the size 4 set are predicted to have equal complexity, which improves the fit of the OQ model.

Lee and Navarro (2002) used a different experimental paradigm to assess the complexity of the size 3 concepts and found that type 1 was easiest to learn, type 3 was second easiest, and that types 2 and 4 were the most difficult to learn. This ordering differs from Aitkin and Feldman's (2006) result in Figure 8 and illustrates that different measures of conceptual complexity do not always produce identical results. One important difference between the two experiments is that Aitkin and Feldman presented all items in the domain simultaneously along with their category labels, whereas Lee and Navarro required participants to learn from feedback as they classified items one by one. Both paradigms have been previously explored (Feldman, 2000; Nosofsky, Gluck, et al., 1994; Shepard et al., 1961), and Shepard et al. (1961) found that they produce consistent measures of conceptual complexity in domains such as (1O,3SF) and (3O,3SF) in Table 2. It is not clear why the two paradigms produce divergent results for the domain with two ternary features, but this result nevertheless exposes a limitation of the OQ model. The OQ model does not explain why different tasks might produce different measures of conceptual complexity, and results of this kind confirm that the representational assumptions of the OQ model will ultimately need to be supplemented with some detailed assumptions about cognitive processing.

Taken overall, the results in Figure 8 provide some initial evidence that language OQ can help to explain how humans think about domains where features take more than two values. The conceptual universe includes many such domains, and existing studies have only explored a small fraction of the full set of possibilities. Ultimately, however, studies involving features with many values may prove to be just as informative about human learning as studies that focus on binary features.

## Relational Domains Related to Shepard et al. (1961)

Previous sections have focused on domains that include features but not relations. I now turn to domains that include relations rather than features. The most natural starting point is domain (3O,1AR) in Table 2, which includes eight items and a single undirected relation and is closely related to the eight-item domains studied in Experiment 1. Domain (3O,1AR) includes the 10 concept types in Figure 2d, and minimal representations for each type are shown in Table 7. As before, these representations can be treated as summaries of more complete descriptions, and the fourth row in Table 3 shows one example of a complete description for this domain. The descriptions assume that relation R cannot meaningfully hold between any object and itself—for example, $R_{aa}$ is not a valid literal. The meaning of any quantified statement should be adjusted accordingly. For example, $\forall_x R_{xa}$ is true in a domain with three objects if and only if $R_{ba}$ and $R_{ca}$ are true.

Complexity values for the 10 types are shown in Table 7. Recall that each literal involving a relation is assigned a complexity value of two. For example, the single literal $R_{ac}$ in the description of concept 1 specifies a relationship between two objects (a and c) and is therefore considered twice as complex as the literal $F_b$, which specifies information about a single object only. The relative complexity values for the 10 concepts are broadly similar to the relative complexities for domain (3O,1AF), which are reproduced in Table 7 for comparison. For example, concept 5 receives the equal lowest complexity score in both domains.

Table 7
*Minimal Descriptions and OQ-Complexities for All 10 Types in Domains (3O,1AR) and (3O,1AF)*

| Type | (3O, 1AR) | *C* | (3O, 1AF) | *C* |
|---|---|---|---|---|
| 1 | $R_{ac}$ | 2 (2) | $F_b$ | 1 (1) |
| 2 | $(\forall_x R'_{xb}) + (\forall_x R_{xb})$ | 4 (4) | $F'_a F'_c + F_a F_c$ | 4 (4) |
| 3 | $R'_{ab} R_{ac} + (\forall_x R_{xb})$ | 6 (6) | $F'_a F_b + F_a F_c$ | 4 (4) |
| 4 | $\forall_x R'_{xc} + (\forall_x R_{xb})$ | 4 (4) | $F'_b F'_c + F_a F_c$ | 4 (4) |
| 5 | $\exists_x \forall_y R_{xy}$ | 2 (2) | $\forall_x \exists_y F_y$ | 1 (1) |
| 6 | $(\forall_x \forall_y R_{xy}) + R'_{ac}(\exists_x R_{xb})$ | 6 (6) | $(\forall_x F_x) + F'_b \exists_y F_y$ | 3 (3) |
| 7 | $\forall_x R_{xc} + (\exists_x R_{xb})(\exists_y \forall_z R'_{yz})$ | 6 (6) | $(\forall_x F_x) + F'_a F_c + F_a(\exists_x \forall_y F'_y)$ | 5 (5) |
| 8 | $(\forall_x \forall_y R_{xy}) + R'_{ac}(\exists_x R'_{xb})$ | 6 (6) | $(\forall_x F_x) + F'_b(\exists_x \forall_y F'_y)$ | 3 (4) |
| 9 | $(\forall_x \forall_y R'_{xy}) + R_{ac}(\exists_x R_{xb})$ | 6 (6) | $(\forall_x F'_x) + F_b(\exists_x \forall_y F_y)$ | 3 (4) |
| 10 | $(\forall_x \forall_y R_{xy}) + (\exists_w \exists_x R_{wx})(\exists_y \forall_z R'_{yz})$ | 6 (6) | $(\forall_x F_x) + (\exists_y \forall_z F_y F'_z)$ | 3 (3) |

*Note.* The complexity values in parentheses show the complexities of the complements of the 10 concepts in Figure 2d.

Crockett (1982) described a study where participants learned several concepts of size 4 from domain (3O,1AR). The objects in the domain were three people a, b, and c, and the relation indicated which pairs of people liked each other. A social network of this kind is said to be *balanced* if people who like each other share the same opinions about the remaining individuals in the network, and if people who dislike each other have different opinions (Heider, 1958). There are four balanced networks, including the network where all individuals like each other and the three networks where two of the individuals like each other and both dislike the third individual. The set of these four networks corresponds to SHJ type VI, and Crockett was therefore especially interested in how well type VI would be learned within domain (3O,1AR).

Crockett (1982) did not describe his study in full, but the concepts considered included SHJ types I, II, and VI, along with one additional concept that belonged to either type III, IV, or V. Crockett (p. 35) reported that types I, II, and VI were learned "with about equal ease" and "significantly more easily" than the additional concept from type III, IV, or V. Although the details are not provided, the most important qualitative result is that type VI does not emerge as the most difficult concept. Because type VI is traditionally considered to be the most difficult type, Crockett interpreted his result as evidence that social learning relies on a domain-specific principle of structural balance.

The complexity values in Table 7 suggest an alternative interpretation. If quantification is allowed, then type VI (corresponding to type 10 in the table) has a relatively simple representation that can be glossed as "networks with either three links or one link." The OQ model therefore suggests that the relative simplicity of type VI in domain (3O,1AR) may depend on a domain-general ability to construct representations that incorporate quantification or counting. This domain-general account predicts that type VI should be relatively easy to learn in any domain that supports quantification or counting. Domain (3O,1AF) is one example, because type VI in this domain can be described as "items where either three squares have slashes or one square has a slash." Consistent with this prediction, Experiment 1 found that type VI (corresponding to type 10 in Figure 4a) was indeed one of the easiest concepts to learn within domain (3O,1AF).

One limitation of Crockett's (1982) study is that it does not acknowledge the existence of 10 types within domain (3O,1AR). For example, focusing on the six SHJ types leads him to suggest that types III, IV, and V are "nonsense classifications" that cannot "be described simply in a word or phrase" (Crockett, 1982, p. 34). The analysis in Figure 2d challenges this claim. In particular, types 5 and 6 in Table 7 are both instances of SHJ type IV, and type 5 can be concisely described as "all networks with two or more links." Future researchers may therefore find it useful to repeat Crockett's study but to include all 10 types in Table 7.

Here, however, I turn to a second relational domain. One motivation for studying relations is that relational representations can sometimes express information that is difficult or impossible for feature-based representations to capture. It follows that there should be relational domains which produce patterns of learning that have no counterpart in feature-based domains. Domain (3O,1AF) is a natural feature-based analog of the relational domain (3O,1AR). Each object in domain (3O,1AF) could be used to represent a pair of individuals, and the slash feature could indicate whether or not the relation holds between the pair. Table 7 suggests, however, that the relative complexities of the types within domain (3O,1AR) are broadly similar to the relative complexities of the corresponding types within domain (3O,1AF). For example, concept 5 is predicted to be relatively easy to learn within both domains. The next section therefore considers a different relational domain that enables a sharper comparison between relational and feature-based representations.

## Relations and Quantification

Each item in domain (3O,1AR) specifies a relation defined over a single set of objects. Relations, however, can also capture relationships between two sets of objects. For example, consider a set of people and a set of foods, and a relation $R(p, f)$ which is true if person $p$ likes food $f$. This section describes results from a concept learning experiment that explored a simple domain of this kind.

The domain considered included a set $O_1$ of three people and a set $O_2$ of two foods, and is labeled (3O × 2O,1AR) or (3O × 2O) for short. The domain includes a single additive relation $R : O_1 \times$

$O_2 \rightarrow \{0, 1\}$ that indicates whether each person likes each food. The domain includes 64 distinct items, and an example is shown in Figure 9c, where the black horizontal lines indicate that a person does not like a certain food. The domain can be used to formulate concepts that are intrinsically relational—for example, the concept that includes all items where some food is liked by all three people.

Domain (3O × 2O) can be compared with a domain (6O × 1O) where there are six people and only one food. Figure 9d shows an example item from this domain. To ensure that this domain matches domain (3O × 2O) as closely as possible, the item represents a binary relation $R : O_1 \times O_2 \rightarrow \{0, 1\}$ where $O_1$ includes six people, and $O_2$ includes a single food. Note, however, that the second argument of the relation is constant, which means that domain (6O × 1O) could also be characterized using a feature $F : O_1 \rightarrow \{0, 1\}$ that indicates whether each individual in $O_1$ likes fruit.

Domain (6O × 1O) can be used to evaluate two hypotheses about how the items in domain (3O × 2O) are mentally represented. One possibility is that the mental representations involved are intrinsically relational—if so, then there should be some relational concepts that are relatively easy to learn in domain (3O × 2O) but that have no simple counterparts in domain (6O × 1O). A second possibility is that the items in domain (3O × 2O) are represented using a collection of six binary features. For example, the first feature might indicate whether Alf likes fruit, the second might indicate whether Alf likes candy, and so on. If so, then learning patterns for domains (3O × 2O) and (6O × 1O) should be similar, because each item in domain (6O × 1O) can also be represented as a collection of feature values.

Like Experiment 1, Experiment 2 also explores the role of quantification. The experiment includes relational concepts that can be concisely described using quantifiers—for example, the

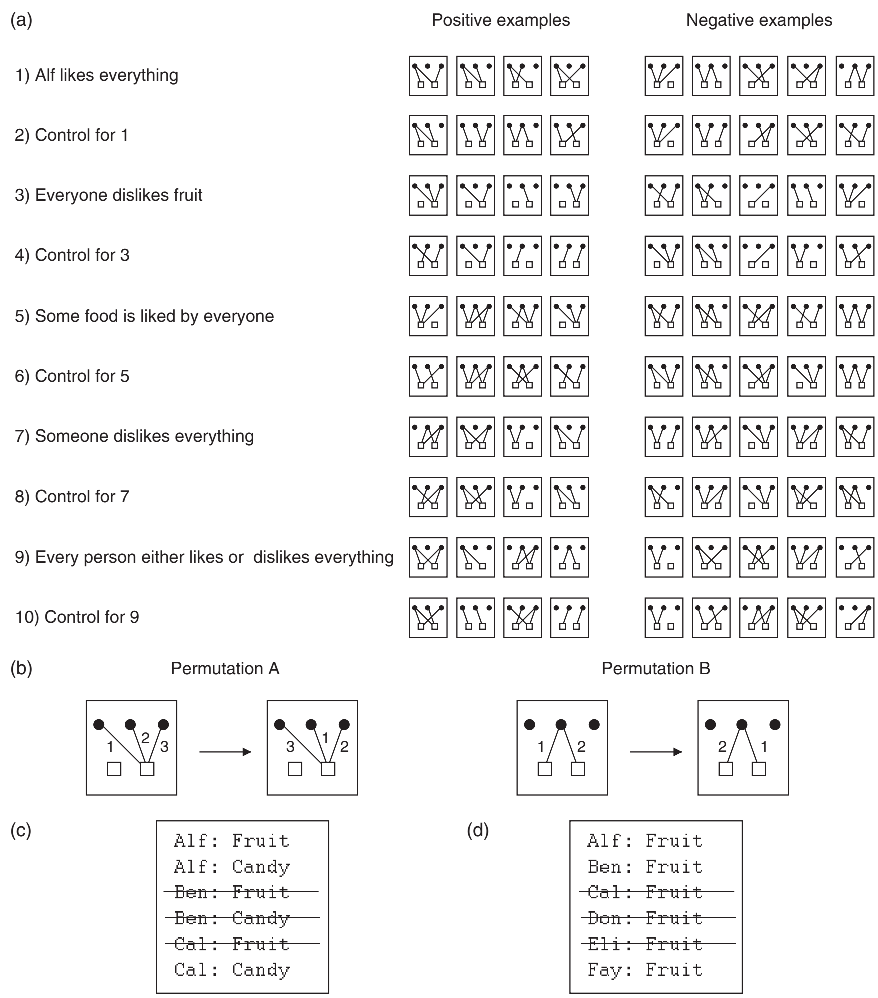



*Figure 9.* (a) Ten concepts explored in Experiment 2. The odd-numbered concepts have concise descriptions that use relations and quantifiers. The even-numbered concepts are controls for the odd-numbered concepts. (b) Permutations used to create the control concepts in (a). Concepts 2, 8, and 10 were created by applying permutation A to concepts 1, 7, and 9, respectively. Concepts 4 and 6 were created by applying permutation B to concepts 3 and 5. (c) An example item from domain (3O × 2O). This item corresponds to the leftmost positive item shown for concept 1 in (a). (d) An example item from domain (6O × 1O) that matches the example item in (c).

concept in domain (3O × 2O) that includes items where *all* of the individuals like fruit. The experiment also includes control concepts that admit no simple relational description, and I predict that the relational concepts will be easier for participants to learn than the control concepts.

**Method.**

***Participants.*** Forty CMU undergraduates participated for course credit.

***Materials.*** Because there are 64 items in each domain the space of all possible concepts is too large to explore in full. Experiment 2 focuses instead on the 10 concepts shown in Figure 9. The items used in the actual experiment are similar to the cards in Figures 9c and 9d, but Figure 9a shows each item as a graph where the three black circles represent people and the two white squares represent foods. Four positive examples and five negative examples of each concept are shown. Because there are 64 items in total, the combined number of positive and negative instances should equal 64 for each concept, but in order to make the task experimentally tractable participants were given only the nine examples shown in Figure 9a.

The 10 concepts in Figure 9a were chosen as follows. The odd-numbered concepts have simple descriptions that involve quantification. For example, concept 1 includes items where Alf likes all foods, and concept 3 includes items where all of the people dislike fruit. The even-numbered concepts are controls for the odd-numbered concepts. Each item specifies binary values for six (person,food) pairs, and the control concepts are created by permuting the order in which these values appear. Figure 9b shows the two permutations that were applied.

***Model predictions.*** Table 8 shows the complexity values for each concept predicted by the OQ model. Complexity values for the complements of these concepts are shown in parentheses. Because the experiment used arbitrary labels (i.e., "red items" and "blue items") for positive and negative examples, the complexities for concepts and their complements are averaged to generate the final predictions. Table C1 in Appendix C shows that similar predictions are generated if the OQ model is adjusted to acknowledge that participants may have preferred to encode the smaller set of examples.

The unadjusted complexity values are plotted in Figure 10b. The model makes two qualitative predictions. First, there will be a difference in learning patterns across the two domains. Second, within domain (3O × 2O) the model predicts that the complexities of the concepts with simple relational descriptions (odd-numbered concepts) will be lower than the complexities of the corresponding control concepts (even-numbered concepts).

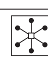
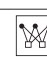
***Procedure.*** Each participant was assigned to domain (3O × 2O) or domain (6O × 1O) and learned the 10 concepts from that domain. The procedure was similar to the procedure for Experiment 1: Participants learned each of the 10 concepts from their assigned domain, and the interface recorded how long they took to learn each concept.

An important departure from Experiment 1 is that the items presented during each test phase were not physically identical to the items presented during the corresponding training phase. Consider domain (3O × 2O) where each item was a card like the example in Figure 9 that contained six (person, food) pairs. During any given phase, the six pairs were listed in the same order on all of the cards. This order, however, differed across training and test phases. Permuting the order in this way ensured that participants could not learn the 10 concepts by focusing on the visual gestalts created by the dark strikethrough bars on each card. Successful completion of the task therefore required participants to focus on the information conveyed by each card, not just its superficial visual appearance. As for all other aspects of the task that were randomized, the order of the (person,food) pairs was pseudo-randomized so that each domain (3O × 2O) participant could be paired with a domain (6O × 1O) participant who experienced exactly the same randomization.

**Results.** Domain (3O × 2O) was easier than domain (6O × 1O) overall. On average, domain (3O × 2O) participants took 1,124 s to learn all 10 concepts, but domain (6O × 1O) participants took 1,321 s (standard deviations were 398 s and 404 s, respectively). As for Experiment 1, all subsequent analyses will focus on normalized learning times that sum to 1 for each participant.

Figure 10a shows the normalized learning times for both conditions. The two qualitative predictions identified previously are

Table 8
*Minimal Descriptions and OQ-Complexities for the 10 Concepts in Experiment 2*

| Type | (6O × 1O)  | *C* | (3O × 2O)  | *C* |
|---|---|---|---|---|
| 1 | $R_{am}R_{dm}$ | 4 (4) | $\forall_y R_{ay}$ | 2 (2) |
| 2 | $R_{am}R_{em}$ | 4 (4) | $R_{am}R_{bn}$ | 4 (4) |
| 3 | $R'_{am}R'_{bm}R'_{cm}$ | 6 (6) | $\forall_x R'_{xm}$ | 2 (2) |
| 4 | $R'_{am}R'_{cm}R'_{em}$ | 6 (6) | $R'_{am}R'_{cm}R'_{bn}$ | 6 (6) |
| 5 | $R_{am}R_{cm} + R'_{am}R_{em}$ | 8 (8) | $\exists_y \forall_x R_{xy}$ | 2 (2) |
| 6 | $R_{am}R_{cm} + R'_{am}R_{bm}$ | 8 (8) | $R_{am}R_{cm} + R'_{am}R_{bm}$ | 8 (8) |
| 7 | $R'_{am}R'_{dm} + R'_{bm}R'_{em} + R'_{cm}R'_{fm}$ | 12 (14) | $\exists_x \forall_y R'_{xy}$ | 2 (2) |
| 8 | $R'_{am}R'_{em} + R'_{bm}R'_{fm} + R'_{cm}R'_{dm}$ | 12 (14) | $R'_{bm}(\forall_y \exists_x R'_{xy}) + (\exists_y R'_{ay})(\forall_y \exists_x R'_{xy})$ | 8 (6) |
| 9 | $R'_{bm}R_{fm} + R'_{cm}R_{dm} + R_{bm}R'_{dm}R_{em}$ | 14 (8) | $\forall_x \forall_y \forall_z (R_{xy} + R'_{xz})$ | 4 (4) |
| 10 | $R'_{bm}R_{dm} + R'_{cm}R_{em} + R_{bm}R'_{em}R_{fm}$ | 14 (8) | $R'_{cm}(\forall_z \exists_x \forall_y R'_{yz}) + R_{an}(\forall_z \exists_x \forall_y R_{yz})$ | 8 (8) |

*Note.* For domain (6O × 1O), objects a–f correspond to the six black circles in the visual representation shown, and object m corresponds to the white square. For domain (3O × 2O), objects a–c correspond to the three black circles in the visual representation shown, and objects m and n correspond to the two white squares. Complexity values in parentheses show the complexities of the complements of the 10 concepts.

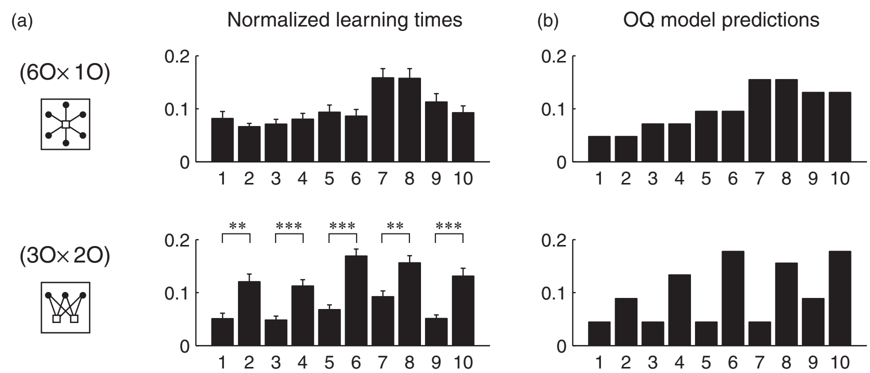



*Figure 10.* (a) Normalized learning times for the 10 concepts in Experiment 2. (b) Normalized OQ-complexity values.

both supported. First, the two domains produce qualitatively different results. Second, the results for domain (3O × 2O) show a sawtooth pattern where each odd-numbered concept is substantially easier than its even-numbered counterpart. Paired-sample *t*-tests indicate that all of these differences are statistically significant at the 0.01 level. These pairwise differences support the hypothesis that the odd-numbered concepts can be concisely represented using quantifiers and relations, but that similarly concise descriptions are not available for the even-numbered concepts. In contrast, the results for domain (6O × 1O) show no systematic differences between odd and even-numbered concepts, and pairwise *t*-tests indicate that none of these differences is statistically significant. The results for domain (6O × 1O) therefore suggest that the results for domain (3O × 2O) depend specifically on the relational structure of this domain.

Figure 11 compares the learning times for the two domains with predictions based on the propositional and the OQ models. Because domain (6O × 1O) has no relational structure, quantification over objects is the only relevant difference between language OQ and propositional logic. It turns out that quantification does not enable concise descriptions of the 10 concepts considered, and the OQ and propositional models make identical predictions about domain (6O × 1O). Figure 11 shows that both models account well for learning times for this domain. In contrast, the OQ model alone accounts for the data from domain (3O × 2O), suggesting that participants think about the concepts in this domain in a way that is intrinsically relational and that incorporates quantification. Each panel in Figure 11 includes a correlation and a bootstrapped 95% confidence interval, and the difference between the correlations achieved by the propositional and OQ models for domain (3O × 2O) is significant at the 0.05 level.

The concept descriptions generated by participants provide further evidence for the role of quantification in concept learning. For example, one description of concept 5 in domain (3O × 2O) indicated that the blue cards are items where "one of the two choices is liked by everyone." One description of concept 7 indicated that the blue cards include items where "there exists someone who doesn't like both fruit and candy." Concept descriptions for domain (6O × 1O) included no comparable descriptions, and many participants indicated that "there was no apparent pattern" and that they simply memorized the groups.

**Discussion.** The results in Figures 10 and 11 support two general conclusions. First, some of the concepts in the conceptual universe are intrinsically relational. The items in domains (3O × 2O) and (6O × 1O) all specify whether six relational links are present or absent, and could therefore be represented as a collection of six binary features. This representation, however, fails to predict the clear difference between patterns of learning across the two domains. The second general conclusion is that the humans rely on quantification when learning relational concepts. The five even-numbered concepts are closely related to the five control concepts, and the only important difference is that the control concepts cannot be concisely described using quantifiers. The even-numbered concepts are consistently easier to learn within domain (3O × 2O), suggesting that participants relied on quantifiers when learning the concepts in this domain.

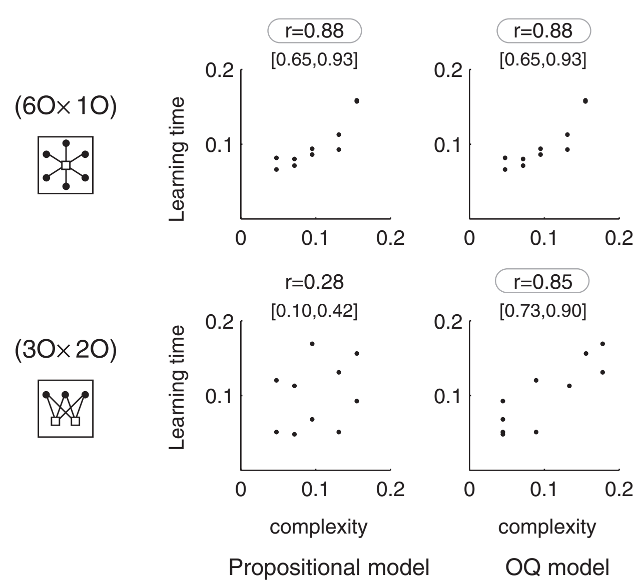



*Figure 11.* Normalized learning times for Experiment 2 plotted against normalized complexity values predicted by the propositional model and the OQ model.

Many previous researchers have studied relational concepts (Doumas et al., 2008; Gentner & Kurtz, 2005; Hummel & Holyoak, 2003; Kemp, Tenenbaum, Niyogi, & Griffiths, 2010; Markman & Stilwell, 2001; Tomlinson & Love, 2007), and it is often suggested that feature-based representations are not rich enough to capture the structure of many concepts. Experiment 2 applied this idea to the study of conceptual complexity. Most previous accounts of conceptual complexity focus on features rather than relations, but Experiment 2 suggests that a comprehensive account of conceptual complexity must incorporate both quantification and relations. The OQ model satisfies this criterion, and accounts relatively well for the results of Experiment 2. This experiment, however, considered just two domains, and additional studies are needed to explore whether the OQ model provides a successful account of conceptual complexity across the many relational domains in the conceptual universe.

## Other Learning Problems

This article has considered how concepts are learned across multiple domains in the conceptual universe. As mentioned earlier, each concept considered can be viewed as a high-level system

$$(O_s = \{s_1, s_2, \ldots, s_n\}, C : O_s \rightarrow \{\text{in, out}\}), \tag{1}$$

where $\{s_1, \ldots, s_n\}$ is the set of items in a given domain, and $C(s_i) =$ in if and only if item $s_i$ belongs to the extension of the concept. Although I have focused on concept learning, other learning problems may be equally informative about the nature of human learning. This section briefly considers two such problems: learning a relation defined over items and learning a single item.

Consider first a problem where participants are asked to learn which pairs of items are valid instances of a relation. For example, the items $\{c_i\}$ might be directed graphs over $n$ nodes (see domain 7 in Table 1), and the relation $R(c_i, c_j)$ might be true whenever graph $c_j$ is the transitive closure of graph $c_i$. Building on the formulation in (1), the relation learning problem can be formalized as the problem of learning a high-level system

$$(O_s = \{s_1, s_2, \ldots, s_n\}, R : O_s \times O_s \rightarrow \{\text{in, out}\}). \tag{2}$$

A closely related problem has been studied in the literature on analogical reasoning. Problems of the form A is to B as C is to ? can be formulated as inductive problems where the task is to learn a binary relation analogous(·,·) given the single example analogous(A, B). This article has focused on problems where the extension of a concept or relation is presented in full and must be memorized, but inductive learning can also be studied across all of the domains in the conceptual universe.

High-level systems like (1) and (2) refer to a set of items $O_s$, where each item may be a semantic system in its own right. This article has focused on problems where the items are taken for granted, and the problem is to partition these items into positive and negative examples of a concept. Learning a single item, however, can also be a challenge, especially in domains where each item carries a substantial amount of information. For example, if each item $c_i$ is a directed graph, memorizing just one of these items may be relatively challenging.

Kemp et al. (2008a) studied the problem of single-item learning in a domain where each item is a directed graph. The six items in their study are shown in Figure 12a. Each graph is defined over a set of objects {a, b, c . . .}, and each panel in Figure 12a shows the edges in a given graph. For example, the star graph is a case where every object sends an edge to object a and all objects have self-edges. The bipartite graph is a case where objects {a, b, c, d, e} all send edges to objects {f, g, h}. The transitive graph is a case where the objects can be organized into a ranking starting with a such that each object sends edges to all lower-ranked objects. All of the graphs have 15 edges in total, and are therefore equally complex from the perspective of a brute-force learner who simply memorizes the list of edges.

Kemp et al. (2008a) measured how long participants took to learn each graph, and asked participants to rate the complexity of each graph after they had learned it. Mean learning times and complexity ratings are shown in Figure 12b. These two measures of subjective complexity produced converging results, and Kemp et al. proposed that the subjective complexity of each graph is captured by its description length in a logical language. The logical descriptions in Figure 12a are minor notational variants of the representations that they proposed, and Figure 12c plots the description length of each graph. Recall that one-place literals (e.g., T(f)) receive a weight of one, and two-place literals (e.g., R(x, y)) receive a weight of two. For example, the description of the bipartite graph has complexity seven because it includes five one-place literals and one two-place literal. Comparing Figures 12b and 12c suggests that the description length measure provides a relatively good account of the relative complexities of the six graphs.

The logical representations in Figure 12a are closely related to the OQ rules considered in previous sections. Each representation in Figure 12a includes multiple components that are separated by periods, and these components correspond to disjuncts in a longer logical statement expressed in disjunctive normal form. The top row of Table 9 shows two rules in disjunctive normal form. Each rule includes three disjuncts, and the second row shows how each disjunct can be converted into an implication with a conjunction on the right hand side.[1] In both cases, the collection of implications is equivalent to the Disjunctive Normal Form (DNF) rule if we assume that the extension specified by the set of implications is the minimal extension consistent with the full set (Muggleton & De Raedt, 1994). For example, we need to assume that an item $i$ belongs to concept $C$ only if it is picked out by at least one of the implications at level 2 of Table 9. The third row of Table 9 shows that a set of implications can be represented in summary format, and these summary representations correspond to the logical representations shown in Figure 12a. For example, the right column of Table 9 suggests how the representation of the transitive graph in Figure 12a can be viewed as a summary of a rule in disjunctive normal form. This representation indicates that ab, bc, cd, de, and ef are all edges in the graph, and that any other pair xz corresponds to an edge if there is some object y such that xy and yz are both edges.

One additional piece of information is required to interpret the logical representations in Figure 12a. Two of the representations include a horizontal line, and the implications below the

[1] Note that the implications take the same general form as the implications that are the foundation of Feldman's (2006) algebraic complexity model.

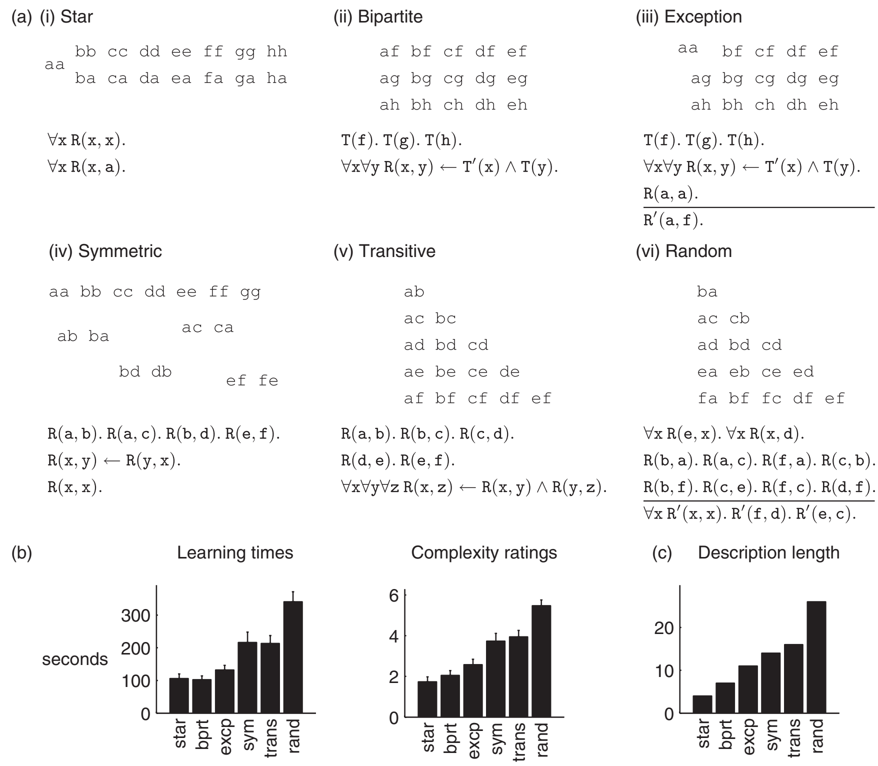



*Figure 12.* Single item learning. (a) Six graphs studied by Kemp et al. (2008a) and minimal descriptions for each graph. (b) Learning times and human complexity ratings for each graph. (c) Description lengths for each graph.

line specify exceptions to the relation defined above the line. For example, the representation of the exception graph indicates that objects f, g, and h belong to the extension of concept T. By the minimal extension assumption, all remaining objects do not belong to the extension of T. The rule R(x, y) ←T′(x) ∧ T(y) indicates that relation R holds between all pairs where x does not belong to T and y belongs to T. The pair af is an exception to this rule, and is therefore specified below the line. A more formal description of how exceptions are represented is provided by Kemp et al. (2008a). Earlier sections of this article did not consider exceptions to logical rules, but previous researchers have found that taking exceptions into account can allow rule-based accounts to provide a closer account of human learning. For example, the RULEX model achieves good fits to human data by combining rules with exceptions (Nosofsky, Palmeri, & McKinley, 1994).

Table 9
*Relationship Between Rules in Disjunctive Normal Form (DNF) and Collections of Implications*

| Representation | $C : O \rightarrow \{0, 1\}$ | $R : O \times O \rightarrow \{0, 1\}$ |
|---|---|---|
| 1. DNF | $\forall i\ C(i) \leftrightarrow (F(i) = 1 \wedge G(i) = 1) \vee$ | $\forall x \forall z\ R(x, z) \leftrightarrow (x = a \wedge z = b) \vee$ |
| | $(F(i) = 0 \wedge G(i) = 0) \vee$ | $(x = b \wedge z = c) \vee$ |
| | $(F(i) = 0 \wedge H(i) = 1).$ | $(\exists y\ R(x, y) = 1 \wedge R(y, z) = 1).$ |
| 2. Implications | $\forall i\ C(i) \leftarrow (F(i) = 1 \wedge G(i) = 1).$ | $\forall x \forall z\ R(x, z) \leftarrow (x = a \wedge z = b).$ |
| | $\forall i\ C(i) \leftarrow (F(i) = 0 \wedge G(i) = 0).$ | $\forall x \forall z\ R(x, z) \leftarrow (x = b \wedge z = c).$ |
| | $\forall i\ C(i) \leftarrow (F(i) = 0 \wedge H(i) = 1).$ | $\forall x \forall z\ R(x, z) \leftarrow (\exists y\ R(x, y) = 1 \wedge R(y, z) = 1).$ |
| 3. Summary | FG. | $R(a, b).$ |
| | $F'G'.$ | $R(b, c).$ |
| | $F'H.$ | $\forall x \forall z\ R(x, z) \leftarrow \exists y\ R(x, y) \wedge R(y, z).$ |

*Note.* Row 1 shows rules in DNF that characterize a concept *C* and a relation *R*. Row 2 shows how these rules can be converted into collections of implications, one for each disjunct. The implications in row 2 are equivalent to the rules in row 1 if we consider the minimal extension consistent with each set of implications. Row 3 shows summary representations of the implications in row 2. The final statement in the summary representation on the right is equivalent to ∀x∀y∀z R(x, z) ← R(x, y) ∧ R(y, z), which appears in Figure 12a.

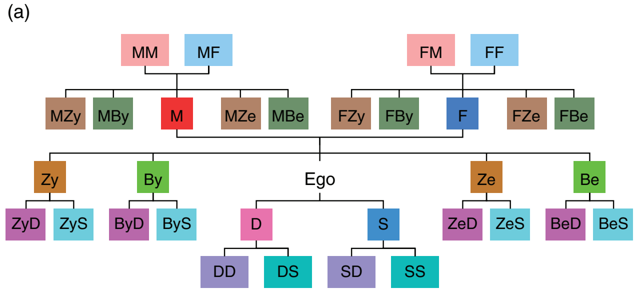
This article has focused on concept learning rather than relation learning or single-item learning, but all three problems provide a way to explore how humans learn and represent information. The ultimate goal is to develop accounts that help to explain how humans solve all of these problems across all of the domains in the conceptual universe. Much work remains to be done, but the results reported in this article provide some initial evidence that the OQ model represents a step in the right direction.

## A Real-World Domain: Kinsfolk

Most of the domains shown in Table 1 and discussed in this article include items that are simple and relatively abstract, such as configurations of different geometric shapes. This section briefly describes how the general approach developed in this article can provide insight into kin classification across cultures.

Domain 9 in Table 1 focuses on siblings alone, but real-world kinship systems also include terms for parents, grandparents, aunts, uncles, and many other relatives. Figure 13a shows a family tree that includes 32 relatives of an individual labeled as "Ego." The colors indicate how English kinship terms organize these relatives into categories. For example, the pink squares indicate that Ego uses the same term ("grandmother") to refer to his mother's mother (MM) and his father's mother (FF). For present purposes, suppose that a kinship system is defined as a partition of the 32 relatives in Figure 13a into categories. English speakers rely on the system shown in Figure 13a, but other languages rely on different partitions of the family tree. For example, English includes just two terms for grandparents ("grandmother" and "grandfather"), but Mandarin includes four terms that can be glossed as "maternal grandmother," "maternal grandfather," "paternal grandmother," and "paternal grandfather."

(a)




(b)

$\texttt{mother}(x, y) \leftrightarrow \texttt{PARENT}(x, y) \wedge \texttt{FEMALE}(x)$
$\texttt{father}(x, y) \leftrightarrow \texttt{PARENT}(x, y) \wedge \texttt{MALE}(x)$
$\texttt{daughter}(x, y) \leftrightarrow \texttt{CHILD}(x, y) \wedge \texttt{FEMALE}(x)$
$\texttt{son}(x, y) \leftrightarrow \texttt{CHILD}(x, y) \wedge \texttt{MALE}(x)$
$\texttt{sister}(x, y) \leftrightarrow \exists z\, \texttt{daughter}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{brother}(x, y) \leftrightarrow \exists z\, \texttt{son}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{sibling}(x, y) \leftrightarrow \exists z\, \texttt{CHILD}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{aunt}(x, y) \leftrightarrow \exists z\, \texttt{sister}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{uncle}(x, y) \leftrightarrow \exists z\, \texttt{brother}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{niece}(x, y) \leftrightarrow \exists z\, \texttt{daughter}(x, z) \wedge \texttt{sibling}(z, y)$
$\texttt{nephew}(x, y) \leftrightarrow \exists z\, \texttt{son}(x, z) \wedge \texttt{sibling}(z, y)$
$\texttt{grandmother}(x, y) \leftrightarrow \exists z\, \texttt{mother}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{grandfather}(x, y) \leftrightarrow \exists z\, \texttt{father}(x, z) \wedge \texttt{PARENT}(z, y)$
$\texttt{granddaughter}(x, y) \leftrightarrow \exists z\, \texttt{daughter}(x, z) \wedge \texttt{CHILD}(z, y)$
$\texttt{grandson}(x, y) \leftrightarrow \exists z\, \texttt{son}(x, z) \wedge \texttt{CHILD}(z, y)$

*Figure 13.* Kin classification. (a) A family tree that includes 32 relatives of an individual labeled as Ego. The colors indicate relatives (e.g., mother's mother [MM] and father's mother [FM]) that belong to the same category (e.g., "grandmother") according to the English kinship system. (b) A minimal description of the English kinship system.

Experiment 1 in this article was developed by systematically characterizing all concepts that can be formulated within four domains, then measuring the subjective complexity of each of these concepts. In principle, the same approach could be applied to the domain of kinship. Because there are 32 relatives in Figure 13a, the number of possible partitions of these relatives is the 32nd Bell number, which is around $10^{26}$. The full space of systems is too large to explore in practice, but an experiment where participants learn different systems from this space could help to reveal the factors that make kinship systems difficult or easy to learn. An alternative approach is to treat kin terminologies across languages as the outcome of a natural experiment (Nerlove & Romney, 1967). Other things being equal, kinship systems settled on by the languages of the world are likely to be subjectively simpler than other systems which are theoretically possible but never found in practice.

The description length hypothesis generates testable predictions about which kinship systems will be observed across languages. Figure 13b shows a logical representation of the English kinship system shown in 13a. The language includes two primitive features (FEMALE(·) and MALE(·)) and two primitive relations (PARENT(·) and CHILD(·)), and all other relations are defined as combinations of these primitives. The system in Figure 13b corresponds to a series of implications—for example, x is the mother of y if and only if x is a parent of y and x is female. The right hand side of each implication is in disjunctive normal form, and the language is therefore consistent with the representation language considered in previous sections of this article.

The complexity of a kinship system can be defined as the length of its shortest logical representation. Kemp and Regier (2012) have explored the idea that this complexity measure can help to predict which kinship systems are found across languages. Complexity is not the only relevant factor. For example, one of the simplest possible kinship systems includes just two terms, one for male relatives and one for female relatives. No language includes a kinship system of this kind, and one possible explanation is that this system is not useful for referring to specific individuals. Kemp and Regier formalized this notion of utility and showed that the kinship systems of the world tend to achieve a near-optimal tradeoff between complexity as measured by description length and utility. For present purposes, the relevant aspect of this result is that the description length hypothesis can help to account for data from an important real-world domain.

## General Discussion

This article introduced a systematic way to characterize a large set of domains and to identify the qualitatively different concepts that can be formulated within these domains. Characterizing the conceptual universe provides a foundation for studies of concept learning, and I discussed empirical studies including two new experiments that explore a total of 11 different domains. Comparing patterns of learning across multiple domains provides strong constraints on theories of concept learning, and should ultimately help to develop general-purpose approaches that help to explain how humans learn and think about all of the domains in the conceptual universe.

As a first step toward a general-purpose model of concept learning, this article proposed and evaluated the OQ model, a rule-based approach that relies on predicate logic as a representation language. Rule-based approaches can be directly applied across all of the domains in the conceptual universe, but it is less clear how alternative approaches to concept learning would achieve the same explanatory scope. I evaluated the OQ model using empirical results from 11 different domains, and the model performed relatively well in all cases. Previous rule-based approaches have focused on propositional logic, but the results of the two experiments suggest that rule-based approaches will need to move beyond propositional logic in order to account for learning across the entire conceptual universe. In particular, the propositional model was unable to explain why participants found it relatively easy to learn concepts that can be concisely described using quantifiers.

The experimental results also suggested that quantification is not indiscriminately applied by human learners. The OQ, FQ, and OQ + FQ models are all very similar, and the only difference between the three is whether they allow quantification over objects, features, or both. The OQ model accounted best for the data, suggesting that humans find it more natural to quantify over objects than features.

The OQ model and the characterization of the conceptual universe provided here should both be viewed as starting points that can be improved upon in subsequent research. The following sections describe some limitations of the current proposals and identify some of the most important directions for future work.

## Toward a More Complete Characterization of the Conceptual Universe

The set-theoretic characterization of the conceptual universe developed here includes a vast number of domains but does not fully cover the space of possibilities. For example, it is straightforward to extend the characterization to include continuous-valued features such as hue, and categorical features that have an ordered set of values (e.g., a size feature that takes values small, medium, and large). More challenging is to allow for interactions between features—for example, if objects can vary with respect to several continuous features, these features may combine in a way that is integral or separable (Garner, 1976; Shepard, 1964).

A second important direction for future work is to provide a systematic characterization of the kinds of problems that can be formulated within each domain. This article focused on a single, simple problem—learning a concept when the concept's extension is presented in full. Other tasks are possible (Goodwin, 2006; Love, 2002; Love et al., 2004; Shepard et al., 1961), including tasks that focus on inductive learning. To understand the relationships between the many possible tasks it may be useful to organize them into a taxonomy. Kemp and Jern (2009b) have taken an initial step in this direction by developing a taxonomy of inductive problems that includes generalization, identification, and categorization, along with several other problems. Combining a taxonomy of problems with a characterization of the conceptual universe may help to identify the most productive directions for future studies to explore. The ultimate goal is clear: psychologists should eventually aim to understand how all of the problems are addressed across all of the domains in the conceptual universe.

## Detecting Objects, Features, and Relations

The OQ model works with items that are characterized as systems of objects, features, and relations. For example, Table 2 indicates that each item in domain (3O,3AF) is characterized as a system that includes three objects, each of which can have one additive feature. Characterizing domain (3O,3AF) in this way seems relatively natural, but note that other characterizations are possible in principle. For example, feature $F_2$ in Table 2 indicates whether or not the second object in an item has spots, but this feature could be replaced with 16 separate features, one for each of the 16 spots. Each item could also be characterized as a system that includes a single object—for example, each item could be an aerial view of a single aquatic creature with head and tail submerged and three square humps protruding above the surface of the water.

These alternative characterizations of domain (3O,3AF) may seem relatively unnatural, but they serve to demonstrate that the OQ model does not work directly with the physical stimuli provided to participants. Instead, the input needed by the model must be supplied by psychological processes that detect the objects, features, and relations that are present in a given stimulus. For simple geometric stimuli like the examples in Table 2, the objects, features, and relations are likely to be supplied by the visual system, and the principles at work may include Gestalt grouping principles. In other cases, however, conceptual knowledge may play a role—for example, understanding that a whale has the feature "breathes air" may depend on prior knowledge that whales are mammals and that mammals breathe air.

Virtually every model of concept learning takes objects and either features, relations or both as input, and questions about the origin of these conceptual elements are raised by all of these models. One such question, however, is especially relevant to this article, and concerns the division of labor between the feature-detection system and the concept learning system. The OQ model is motivated by the idea that expressive representation languages are needed to explain how humans construct concepts that depend on quantification and relations. An alternative approach might propose that the representation language used for concept learning is very simple, but that representations in this language make use of emergent features including relational features and features that depend on quantification.[2] For example, consider the stimuli used for domain (3O,1AF) in Experiment 1. If "has exactly one slash" and "has exactly three slashes" are among the features available, then concept 10 in Table 4 can be captured using a simple representation language such as Boolean logic that does not incorporate quantification or counting. Supporters of this view may wish to consider a variant of the OQ model where language OQ is used to explain how emergent features are constructed from simpler components, and where an existing feature-based approach such as propositional logic or a connectionist network is used to explain how concepts are learned given these features. Note, however, that language OQ still plays a critical role in this model, and cannot be removed without substituting some alternative account of how relational features and features that depend on quantification are constructed.

[2] I thank an anonymous reviewer for proposing this account.

## The Minimum Description Length Approach

As evaluated in this article, the OQ model focuses on the problem of memorizing a set of positive and negative examples. The model, however, can be viewed as a special case of the minimum description length (MDL) approach to learning and inference (Grünwald, 2007), and the literature on the MDL principle illustrates how the core ideas behind the OQ model can be applied to a range of different problems. This section explains how the MDL approach allows the OQ model to be formulated in more general terms, and shows how this more general formulation explains why concept learners sometimes acquire rules that are longer than the shortest rule that correctly classifies the observed examples.

Suppose that a learner observes a data set $d$ and attempts to find the shortest description of the data. I assume that this description makes use of a rule $r$ such that the total description length is

$$L(d|r) + L(r), \tag{3}$$

where $L(r)$ is the description length of rule $r$ and $L(d|r)$ is the description length of the data given rule $r$. The same idea can be formulated using probability theory. If the description length $L(x)$ is equal to $-\log(P(x))$, which means that the description length of an observation is inversely related to its probability, then the rule that minimizes Equation 3 will be the same as the rule that maximizes the posterior probability $P(r|d) \propto P(d|r)P(r)$ (Grünwald, 2007). Although probabilistic inference and the MDL approach are equivalent for some purposes, the MDL approach is especially natural when considering issues related to mental representation, which is why this article focused on description lengths $L(r)$ rather than prior probabilities $P(r)$.

For the canonical problem considered in this article, the data set $d$ consists of a set of stimuli $s$ along with category labels $c$ for each stimulus. Equation 3 then becomes

$$L(c, s|r) + L(r) = L(c|s, r) + L(s|r) + L(r) \tag{4}$$

$$= L(s|c, r) + L(c|r) + L(r). \tag{5}$$

Equations 4 and 5 are equivalent but Equation 4 is most useful when thinking about the tasks considered in this article. For example, consider the paradigm used in Experiment 1, where the stimuli $s$ include all items in the domain. Because the stimuli are chosen in a way that does not depend on $r$, $L(s|r) = L(s) = k_1$ where $k_1$ is a constant that does not depend on $r$. $L(c|s, r) = 0$, because rule $r$ must perfectly classify the stimuli in $s$, which means that the category labels $c$ carry no new information given $s$ and $r$. As a result, the total description length is equal to $k_1 + L(r)$, and this sum is minimized by choosing the shortest rule $r$ that correctly classifies all of the observed stimuli.

Although the analysis just described predicts that participants in Experiment 1 will tend to learn the shortest rule that accounts for the observed examples, the MDL approach predicts that people will learn non-minimal rules in other contexts. Consider, for example, the problem of learning from positive examples (Chater & Vitanyi, 2007; Conklin & Witten, 1994; Hsu & Chater, 2010; Tenenbaum & Griffiths, 2001). Given positive examples only, the simplest consistent rule always states that all items are examples of the concept, but in many cases people infer a more specific rule. For example, suppose that a learner observes four positive examples which are brown squares of different sizes. It is possible that the underlying category includes all squares, or all brown things, but the more intuitive conclusion is that the category includes all brown squares.

The MDL approach predicts that whether or not people infer minimal rules will depend on assumptions about how the stimuli were sampled. Previous studies have explored different sampling assumptions (Navarro, Dry, & Lee, 2012; Tenenbaum & Griffiths, 2001) and have provided some evidence that people's inferences are sensitive to these assumptions (Xu & Tenenbaum, 2007). Equation 5 explains why people infer non-minimal rules when the stimuli $s$ are assumed to be drawn only from the set of positive examples. If the stimuli are guaranteed to be positive examples, then $c$ will always include positive labels regardless of the underlying rule $r$. As a result, rule $r$ is uninformative about $c$ and $L(c|r) = L(c) = k_2$ where $k_2$ is a constant that is independent of $r$. The total description length in Equation 5 is therefore equal to $L(s|c, r) + k_2 + L(r)$. In general, there will be a tradeoff between the first and final terms in the sum. Consider again the example where a learner observes four brown squares. If rule $r_1$ indicates that all members of the category are brown and square, then $L(r_1)$ will be relatively high but $L(s|c, r_1)$ will be relatively low because the description of $s$ does not need to specify that each stimulus is brown and square. On the other hand, if rule $r_2$ specifies only that members of the category are square, then $L(r_2)$ will be relatively low but $L(s|c, r_2)$ will be relatively high because the description of each stimulus will need to specify that it is brown. As this example suggests, the tradeoff between $L(s|c, r)$ and $L(r)$ means that the rule that minimizes the sum of the two will not always be the same as the simplest rule that correctly classifies the observed examples.

A probabilistic approach equivalent to the MDL approach just described has been used to model experiments where learners acquire concepts from positive examples (Kemp et al., 2008a; Kemp & Jern, 2009a). Future studies can consider whether alternative sampling assumptions will allow the MDL approach to account for other cases where humans learn non-minimal rules (Medin, Wattenmaker, & Michalski, 1987; Nosofsky, 1991; Nosofsky, Clark, & Shin, 1989). For example, Medin et al. (1987) considered a task where participants were shown positive and negative examples of a concept, and found that participants often inferred non-minimal rules in this setting. The MDL approach may be able to account for this result provided that the stimuli $s$ are assumed to be chosen in a way that depends on the underlying rule. One possible assumption is that the stimuli were sampled in a way that guarantees equal numbers of positive and negative examples. Sampling stimuli in this way ensures that $L(c|r)$ in Equation 5 depends on $r$, which opens up the possibility that a rule can minimize the total description length in Equation 5 even though it does not minimize $L(r)$.

## Compositionality and Natural Language

This article is motivated in part by the idea that compositional representation languages generate large sets of structures that can be used to capture the meanings of natural language expressions (Jackendoff, 1983; Larson & Segal, 1995; Montague, 1973). For example, I suggested that the compositional nature of predicate logic allows it to capture the meanings of natural language phrases such as "brown square" and "armed robbery." Predicate logic also provides a promising way to capture the meanings of goal-derived categories that are typically expressed using more complex phrases such as "things to take on a camping trip" (Barsalou, 1983). Note, however, that this article has not provided a theory that can be used to map natural language phrases into predicate logic structures. Words can combine in many different ways, and developing a general theory that picks out the predicate logic structure that corresponds to a given phrase is a major challenge.

The literature on noun-noun combinations illustrates some of the issues that arise when explaining how people interpret the meanings of phrases (Costello & Keane, 2000; Hampton, 1997; Murphy, 1990; Wisniewski, 1997). Some noun-noun combinations appear to correspond to simple conjunctions. For example, "orphan girl" refers to any person who is an orphan and a girl. The meaning of this phrase can therefore be captured by using a conjunction to combine predicate logic structures that correspond to "orphan" and "girl." Many combinations, however, do not correspond to conjunctions. For example, a "forest walk" is not a forest that is also a walk, but rather a walk that takes place in a forest. The theoretical approach developed in this article proposes that the meanings of "orphan girl," "forest walk," and other phrases can be captured using predicate logic, but the ways in which predicate logic structures for individual words combine to form structures for phrases may differ from case to case.

The considerations just described open up a research program on conceptual combination that has been pursued by Costello and Keane (2000) among others. These researchers have developed a computational model of conceptual combination which uses predicate logic to represent both the meanings of individual nouns and the meanings of noun-noun compounds. The model takes predicate logic structures for two nouns as input, and generates as output a structure that captures the meaning of the corresponding noun-noun compound. The model of Costello and Keane captures some of the principles that humans use to interpret novel phrases, but developing a fully general account of natural language interpretation is a long-term research challenge.

## Process Models of Concept Learning

The OQ model is primarily intended to address the question of how concepts are represented, and the core component of the model is representation language OQ. Many researchers have pointed out, however, that mental representations and mental processes cannot be studied separately, and that every proposal about mental representation should be accompanied by a proposal about the processes that operate over that representation (Anderson, 1978). This article has relied on a simple hypothesis about mental processes— the hypothesis that these processes are sensitive to the length of a description formulated in a compositional language. Previous studies have applied the same basic hypothesis to problems from a range of cognitive domains, including language learning (Chater & Vitanyi, 2007; Chomsky, 1975), visual perception (Chater, 1996; Leeuwenberg, 1971), sequence learning (Simon, 1972), similarity judgment (Hahn, Chater, & Richardson, 2003), generalization (Chater & Vitanyi, 2003a), and concept learning (Feldman, 2000; Kemp et al., 2008a; Medin et al., 1987; Pothos & Chater, 2002).

As mentioned earlier, the description length hypothesis is appealing in part because it provides an idealized account of processing that abstracts away most of the details. The hypothesis can therefore be used as a simple initial strategy for evaluating claims about mental representation. This article used the hypothesis to assess the psychological merits of four languages: OQ, FQ, OQ + FQ, and propositional logic. Given that the OQ model appears to account relatively well for data from a broad range of domains, future research can explore process models that combine the representational assumptions of the OQ model with more detailed processing assumptions. Previous researchers have developed process models that help to explain how propositional rules are learned and used (Bradmetz & Mathy, 2008; Fific et al., 2010; Fific, Nosofsky, & Townsend, 2008; Little, Nosofsky, & Denton, 2011; Nosofsky, Palmeri, & McKinley, 1994), and similar approaches may help to explain how representations in language OQ are learned.

Developing process models that incorporate representations in language OQ should help to overcome several limitations of the current approach. The OQ model accounts relatively well for average responses across participants, but is not designed to account for individual differences. The adjusted OQ model proposes that participants may identify two kinds of rules: some participants identify the simplest description of a set of examples, but others rely on a rule that simply enumerates the examples observed. In reality, however, there are more than two possible rules that might be considered. The adjusted OQ model could be used as the starting point for a richer process model which assumes that participants select among multiple rules, where the probability of choosing a given rule depends in part on its description length. Previous rule-based approaches have been able to account for data at the individual level (Goodman et al., 2008; Nosofsky, Palmeri, & McKinley, 1994), and it should be possible to develop models that combine language OQ with the processing assumptions of previous rule-based approaches.

Appropriate processing assumptions may also allow the OQ model to account for the graded nature of concepts. At first it may seem that models that rely on rules have no way to account for typicality effects and graded category membership. These phenomena, however, can be captured by rule-based accounts that generate multiple rules for each concept (Goodman et al., 2008). Borderline examples of the concept may satisfy only one of the rules, but the most typical members of the concept will satisfy all of the rules. Goodman et al. (2008) showed that this general approach can be captured by working with probability distributions over rules, and described a probabilistic rule-based approach that is able to account for prototype and typicality effects.

A single set of processing assumptions is unlikely to account for all of the ways in which rules are learned and used, and it may be necessary to develop different process models for different contexts. The OQ model can already explain to some extent why different rules might be learned when the same concept is encountered in different contexts. Consider, for instance, a problem where all members of a concept have three distinctive features. The minimal rule for classifying items as positive or negative examples of the concept might incorporate just one of the three features. A different experimental task, however, might require participants to make inferences about unobserved features of category members (Yamauchi & Markman, 1998). Now the shortest adequate rule will need to include all three of the distinctive features. As this example suggests, the description length hypothesis predicts that learners will tend to construct the simplest concept representation that is consistent with their goals and the task constraints. Formalizing the goals and constraints, however, will often require more detailed processing assumptions.

This section has described many directions in which the OQ model can be extended. The core proposal of the model is that humans construct rules in a compositional language like OQ. This proposal, however, is consistent with a large family of models, including models that learn multiple rules for a given concept, models that learn different rules depending on the context, and models that sample from a probability distribution over rules. Researchers have explored all of these directions using models that rely on propositional logic, and the results in this article suggest that similar investigations using OQ as the underlying representation language may be productive.

## Conclusion

Psychological accounts of concept learning often focus on a single domain or a handful of domains, but should ultimately aim to explain how human learning operates across all of the domains in the conceptual universe. This article has made a start in this direction by providing a formal characterization of a large space of domains and describing a computational model that promises to explain how humans learn concepts across the entire space. The model is founded on two key ideas: first, that humans make use of a compositional representation language that allows them to construct concepts within many different domains, and second, that the psychological complexity of a concept is determined by the length of its representation in this language.

The model was evaluated using data drawn from 11 different domains, including domains considered by previous accounts of Boolean concept learning and previously unstudied domains that emphasize the role of quantification and relations. This collection of 11 domains is relatively large by the standards of previous research, but the conceptual universe contains a vast number of additional domains and many different tasks can be formulated within each of these domains. Exploring these domains and tasks is a long-term challenge that may never be completed in full, but the results generated at each step of the way seem likely to prove informative about the nature of human learning.

## References


Aitkin, C. D., & Feldman, J. (2006). Subjective complexity of categories defined over three-valued features. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 961–966). New York, NY: Psychology Press.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277. doi:10.1037/0033-295X.85.4.249

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11,* 211–227. doi:10.3758/BF03196968

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts* (pp. 21–74). Hillsdale, NJ: Erlbaum.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition, 56,* 1–30. doi:10.1016/0010-0277(94)00648-5

Bourne, L. E. (1970). Knowing and using concepts. *Psychological Review, 77,* 546–556. doi:10.1037/h0030000

Bradmetz, J., & Mathy, F. (2008). Response times seen as decompression times in Boolean concept use. *Psychological Research/Psychologische Forschung, 72,* 211–234. doi:10.1007/s00426-006-0098-7

Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic.* Mahwah, NJ: Erlbaum.

Brooks, P. J., & Braine, M. D. S. (1996). What do children know about the universal quantifiers all and each? *Cognition, 60,* 235–268. doi:10.1016/0010-0277(96)00712-3

Bruner, J. A., Goodnow, J. S., & Austin, G. J. (1956). *A study of thinking.* New York, NY: Wiley.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247–303. doi:10.1207/s15516709cog2303_1

Carey, S. (2009). *The origin of concepts.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195367638.001.0001

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review, 103,* 566–581. doi:10.1037/0033-295X.103.3.566

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52*(A), 273–302. doi:10.1080/027249899391070

Chater, N., & Vitanyi, P. (2003a). The generalized universal law of generalization. *Journal of Mathematical Psychology, 47,* 346–369. doi:10.1016/S0022-2496(03)00013-0

Chater, N., & Vitanyi, P. (2003b). Simplicity: A unifying principle in cognitive science. *Trends in Cognitive Sciences, 7,* 19–22. doi:10.1016/S1364-6613(02)00005-0

Chater, N., & Vitanyi, P. (2007). "Ideal learning" of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology, 51,* 135–163. doi:10.1016/j.jmp.2006.10.002

Chomsky, N. (1975). *The logical structure of linguistic theory.* Chicago, IL: University of Chicago Press.

Conklin, D., & Witten, I. H. (1994). Complexity-based induction. *Machine Learning, 16,* 203–225. doi:10.1007/BF00993307

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science, 24,* 299–349. doi:10.1207/s15516709cog2402_4

Cottrell, N. B. (1975). Heider's structural balance principle as a conceptual rule. *Journal of Personality and Social Psychology, 31,* 713–720. doi:10.1037/0022-3514.31.4.713

Crockett, W. H. (1982). Balance, agreement, and positivity in the cognition of small social structures. In *Advances in Experimental Social Psychology* (Vol. 15, pp. 1–57). doi:10.1016/S0065-2601(08)60294-4

Davidson, D. (1967). The logical form of action sentences. In N. Rescher (Ed.), *The logic of decision and action* (pp. 81–95). Pittsburgh, PA: University of Pittsburgh Press.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115,* 1–43. doi:10.1037/0033-295X.115.1.1

Fass, D., & Feldman, J. (2003). Categorization under complexity: A unified MDL account of human learning of regular and irregular categories. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15* (pp. 34–35). Cambridge, MA: MIT Press.

Feldman, J. (2000, October 5). Minimization of Boolean complexity in human concept learning. *Nature, 407,* 630–633. doi:10.1038/35036586

Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology, 50,* 339–368. doi:10.1016/j.jmp.2006.03.002

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk and decision-bound approaches. *Psychological Review, 117,* 309–348. doi:10.1016/j.jmp.2006.03.002

Fific, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 356–375. doi:10.1037/0096-1523.34.2.356

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York, NY: Holt, Rinehart, and Winston.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8,* 98–123. doi:10.1016/0010-0285(76)90006-2

Garner, W. R. (1978a). Aspects of a stimulus. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (p. 99–133). Hillsdale, NJ: Erlbaum.

Garner, W. R. (1978b). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General, 107,* 287–308. doi:10.1037/0096-3445.107.3.287

Gati, I., & Tversky, A. (1982). Representation of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 325–340. doi:10.1037/0096-1523.8.2.325

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16,* 341–370. doi:10.1016/0010-0285(84)90013-6

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170. doi:10.1207/s15516709cog0702_3

Gentner, D., & Kurtz, K. (2005). Relational categories. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the lab* (pp. 151–175). doi:10.1037/11156-009

Giralt, N., & Bloom, P. (2000). How special are objects? Children's reasoning about objects, parts, and holes. *Psychological Science, 11,* 497–501. doi:10.1111/1467-9280.00295

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32,* 108–154. doi:10.1080/03640210701802071

Goodwin, G. P. (2006). *How individuals learn simple Boolean systems and diagnose their faults* (Unpublished doctoral dissertation). Princeton University, NJ.

Goodwin, G. P., & Johnson-Laird, P. N. (2011). Mental models of Boolean concepts. *Cognitive Psychology, 63,* 34–59. doi:10.1016/j.cogpsych.2011.04.001

Grünwald, P. (2007). *The minimum description length principle.* Cambridge, MA: MIT Press.

Hahn, U., Chater, N., & Richardson, L. B. C. (2003). Similarity as transformation. *Cognition, 87,* 1–32. doi:10.1016/S0010-0277(02)00184-1

Hampton, J. A. (1997). Conceptual combination. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 133–159). Cambridge, MA: MIT Press.

Hayes, P. J. (1978). The naive physics manifesto. In D. Michie (Ed.), *Expert systems in the micro-electronic age* (pp. 242–270). Edinburgh, Scotland: Edinburgh University Press.

Heider, F. (1958). *The psychology of interpersonal relations.* doi:10.1037/10628-000

Hobbs, J. R., & Moore, R. C. (Eds.). (1985). *Formal theories of the commonsense world.* New York, NY: Ablex.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355. doi:10.1207/s15516709cog1303_1

Hsu, A. S., & Chater, N. (2010). The logical problem of language acquisition: A probabilistic perspective. *Cognitive Science, 34,* 972–1016. doi:10.1111/j.1551-6709.2010.01117.x

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264. doi:10.1037/0033-295X.110.2.220

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of discrete entities. *Cognition, 85,* 203–221. doi:10.1016/S0010-0277(02)00088-4

Jackendoff, R. (1983). *Semantics and cognition.* Cambridge, MA: MIT Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2008a). Learning and using relational theories. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in Neural Information Processing Systems 20* (pp. 753–760). Cambridge, MA: MIT Press.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2008b). Theory acquisition and the language of thought. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1606–1611). Austin, TX: Cognitive Science Society.

Kemp, C., & Jern, A. (2009a). Abstraction and relational learning. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 934–942). Retrieved from http://books.nips.cc/nips22/NIPS2009_preface.pdf

Kemp, C., & Jern, A. (2009b). A taxonomy of inductive problems. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 255–260). Austin, TX: Cognitive Science Society.

Kemp, C., & Regier, T. (2012, May 25). Kinship categories across languages reflect general communicative principles. *Science, 336,* 1049–1054. doi:10.1126/science.1218811

Kemp, C., Tenenbaum, J. B., Niyogi, S., & Griffiths, T. L. (2010). A probabilistic model of theory formation. *Cognition, 114,* 165–196. doi:10.1016/j.cognition.2009.09.003

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17,* 248–294. doi:10.1016/0010-0285(85)90009-X

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797–2822. doi:10.2307/1131753

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. doi:10.1037/0033-295X.99.1.22

Lafond, D., Lacouture, Y., & Cohen, A. L. (2009). Decision-tree models of categorization response times, choice proportions, and typicality judgments. *Psychological Review, 116,* 833–855. doi:10.1037/a0017188

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology, 51,* 57–74. doi:10.1016/j.jmp.2006.12.001

Larson, R., & Segal, G. (1995). *Knowledge of meaning: An introduction to semantic theory.* Cambridge, MA: MIT Press.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian methods. *Journal of Mathematical Psychology, 55,* 1–7. doi:10.1016/j.jmp.2010.08.013

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review, 9,* 43–58. doi:10.3758/BF03196256

Leeuwenberg, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology, 84,* 307–349. doi:10.2307/1420464

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface.* Cambridge, MA: MIT Press.

Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1–27. doi:10.1037/a0021330

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review, 9,* 829–835. doi:10.3758/BF03196342

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332. doi:10.1037/0033-295X.111.2.309

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence, 13,* 329–358. doi:10.1080/09528130110100252

Mathy, F. (2010). The long-term effect of relational information in classification learning. *European Journal of Cognitive Psychology, 22,* 360–390. doi:10.1080/09541440902903603

Mathy, F., & Bradmetz, J. (2004). A theory of the graceful complexification of concepts and their learnability. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 22,* 41–82.

Mathy, F., & Bradmetz, J. (2011). An extended study of the nonindependence of stimulus properties in human classification learning. *The Quar-*

terly Journal of Experimental Psychology, 64,* 41–64. doi:10.1080/17470218.2010.493972

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science, 11,* 299–339. doi:10.1207/s15516709cog1103_3

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York, NY: McGraw-Hill.

Montague, R. (1973). The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). doi:10.1007/978-94-010-2506-5_10

Muggleton, S., & De Raedt, L. (1994). Inductive logic programming: Theory and methods. *Journal of Logic Programming, 19–20,* 629–679. doi:10.1016/0743-1066(94)90035-3

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language, 29,* 259–288. doi:10.1016/0749-596X(90)90001-G

Navarro, D. J., Dry, M. J., & Lee, M. D. (2012). Sampling assumptions in inductive generalization. *Cognitive Science, 36,* 187–223. doi:10.1111/j.1551-6709.2011.01212.x

Neisser, U., & Weene, P. (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology, 64,* 640–645. doi:10.1037/h0042549

Nerlove, S. B., & Romney, A. K. (1967). Sibling terminology and cross-sex behavior. *American Anthropologist, 69,* 179–187. doi:10.1525/aa.1967.69.2.02a00050

Newell, A. (1989). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Nosofsky, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition, 19,* 131–150. doi:10.3758/BF03197110

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282–304. doi:10.1037/0278-7393.15.2.282

Nosofsky, R. M., Gluck, M., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22,* 352–369. doi:10.3758/BF03200862

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79. doi:10.1037/0033-295X.101.1.53

Oaksford, M., & Chater, N. (2002). Commonsense reasoning, logic, and human rationality. In R. Elio (Ed.), *Common sense reasoning and rationality* (pp. 174–214). doi:10.1093/0195147669.003.0009

Piantadosi, S. T., Goodman, N. D., & Tenenbaum, J. B. (2010). Beyond Boolean logic: Exploring representation languages for learning complex concepts. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 859–864). Austin, TX: Cognitive Science Society.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26,* 303–343. doi:10.1207/s15516709cog2603_6

Pustejovsky, J. (1991). The syntax of event structure. *Cognition, 41,* 47–81. doi:10.1016/0010-0277(91)90032-Y

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

Rodriguez, P., Wiles, J., & Elman, J. L. (1999). A recurrent neural network that learns to count. *Connection Science, 11,* 5–40. doi:10.1080/095400999116340

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211–236). Orlando, FL: Academic Press.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General, 133,* 534–553. doi:10.1037/0096-3445.133.4.534

Sassoon, G. W. (2011). Adjectival vs. nominal categorization processes. *Belgian Journal of Linguistics, 25,* 104–147. doi:10.1075/bjl.25.06sas

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Shapiro, S. C., & Rapaport, W. J. (1992). The SNePS family. In S. C. Shapiro & W. J. Rapaport (Eds.), *Semantic networks in artificial intelligence* (pp. 243–275). Oxford, England: Pergamon Press.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54–87. doi:10.1016/0022-2496(64)90017-3

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13). (Whole No. 517)

Shipley, E. F., & Shepperson, B. (1990). Countable entities: Developmental changes. *Cognition, 34,* 109–136. doi:10.1016/0010-0277(90)90041-H

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review, 79,* 369–382. doi:10.1037/h0033118

Sowa, J. (1984). *Conceptual structures: Information processing in mind and machine.* Reading, MA: Addison-Wesley.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14,* 29–56. doi:10.1207/s15516709cog1401_3

Stenning, K., & van Lambalgen, M. (Eds.). (2008). *Human reasoning and cognitive science.* Cambridge, MA: MIT Press.

Szymanik, J., & Zajenkowski, M. (2010). Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science, 34,* 521–532. doi:10.1111/j.1551-6709.2009.01078.x

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. doi:10.1017/S0140525X01000061

Thomas, C. (2009). *An exploration of semantic formalisms—Part I: A comparison of first order predicate calculus, intensional logic and conceptual structures* (Tech. Rep. No. 2009-554). Kingston, Ontario, Canada: Queens University, School of Computing.

Tomlinson, M. T., & Love, B. C. (2007). Relation-based categories are easier to learn than feature-based categories. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1557–1562). Austin, TX: Cognitive Science Society.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi:10.1037/0033-295X.84.4.327

Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology, 50,* 501–510. doi:10.1016/j.jmp.2006.05.007

Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology, 53,* 203–221. doi:10.1016/j.jmp.2009.04.009

Vigo, R. (2011). Towards a law of invariance in human conceptual behavior. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2580–2585). Austin, TX: Cognitive Science Society.

Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition, 90,* 163–191. doi:10.1016/S0010-0277(03)00143-4

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review, 4,* 167–183. doi:10.3758/BF03209392

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10,* 288–297. doi:10.1111/j.1467-7687.2007.00590.x

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124–148. doi:10.1006/jmla.1998.2566

(*Appendices follow*)

## Appendix A

## Counting Concept Types

This appendix describes a method for computing the number of qualitatively different concepts that exist within a given domain. The method is based on the approach that Shepard, Hovland, and Jenkins (1961) used to derive the six concept types in domain (1O,3SF) but is formulated in a way that can be applied to any domain in the conceptual universe.

All of the domains in Table 2 are described using symbols for objects (e.g., $o_1$ and $o_2$), features ($F_1$ and $F_2$), feature values ($v_1$ and $v_2$), and relations ($R_1$ and $R_2$). Some aspects of the mapping between these symbols and the items in the domain are arbitrary. In domain (1O,3SF), for example, choosing to label the size feature $F_1$ rather than $F_2$ requires an arbitrary decision, and choosing to label the feature value "small" as $v_1$ rather than $v_2$ requires a second arbitrary decision. Any pair of extensions will be considered instances of the same type if they are identical up to arbitrary decisions of this kind. More formally, any pair of extensions belong to the same type if one can be converted into the other by applying a structure-preserving transformation. The structure-preserving transformations for domain (1O,3SF) are shown in the top row of Table A1. For example, the set $\{F_1, F_2, F_3\}$ indicates that permuting the feature labels does not change the basic structure of a concept. Allowing for the transformations shown in the table reveals that each four-item extension in domain (1O,3SF) corresponds to one of the six concept types shown in Figure 2b.

The remaining rows in Table A1 show structure-preserving transformations that are appropriate for the remaining domains in Table 2. In domain (3O,1SF), the only acceptable transformations are those that permute the labels of the objects and the values of the single feature *F*. Allowing for these transformations produces the nine types shown in Figure 2c. In Figure 2c, type 5 could be converted into type 6 by inverting the third binary variable so that values 0 and 1, respectively, indicate that the third object has values $v_2$ and $v_1$ for feature *F*. A transformation of this kind explains why types 5 and 6 are equivalent within domain (3O,3SF), but this transformation is no longer acceptable within domain (3O,1SF). Because the same feature applies to all three objects, feature values $v_1$ and $v_2$ cannot be exchanged for just one of the objects in isolation. Instead, any transformation of this kind must be applied to all three objects uniformly.

The final two columns in Table A1 show the number of types of size four that emerge when the transformations listed in the table are taken into account. The second-last column shows the number of types when the symbols that indicate whether an item belongs to a concept's extension can be exchanged. In Experiment 1, participants were asked to learn an extension of size 4 by sorting eight items into a "blue group" of size four and a "red group" of size four. Labeling the items that belong to the extension as red rather than blue requires an arbitrary decision, and the basic structure of a concept is therefore preserved by a transformation that exchanges the roles of red and blue. If the two groups are labeled as "blickets" and "non-blickets," then these labels are no longer exchangeable, because there is now a qualitative difference between the extension of a concept and its complement. The final column of Table A1 shows the number of concept types when the labels for the extension and its complement cannot be exchanged. In the three domains that are well-characterized by the SHJ types, the final two columns are identical. In domain (3O,1SF), however, the number of types increases from 9 to 12 if extensions and complements cannot be exchanged. Note, for example, that type 2 in Figure 2c is qualitatively different from its complement. In particular, type 2 includes both items where all three objects take identical values for *F*, and its complement includes neither of these items. Type 1, however, belongs to the same type as its complement, and exchanging the Boolean codes used in Figure 2c for $v_1$ and $v_2$ provides a way to transform one into the other.

Table A1
*Number of Four-Item Types Within the Domains in Table 2*

| Domain | Structure preserving transformations | No. of types if {in, out} | No. of types if {in}, {out} |
|---|---|---|---|
| 1. (1O,3SF) | $\{F_1, F_2, F_3\}$, $\{v_1, v_2\}$, $\{v_3, v_4\}$, $\{v_5, v_6\}$ | 6 | 6 |
| 2. (1O,3AF) | $\{F_1, F_2, F_3\}$ | 10 | 20 |
| 3. (3O,3SF) | $\{o_1, o_2, o_3\}$, $\{F_1, F_2, F_3\}$, $\{v_1, v_2\}$, $\{v_3, v_4\}$, $\{v_5, v_6\}$ | 6 | 6 |
| 4. (3O,3AF) | $\{o_1, o_2, o_3\}$, $\{F_1, F_2, F_3\}$ | 10 | 20 |
| 5. (3O,1SF) | $\{o_1, o_2, o_3\}$, $\{v_1, v_2\}$ | 9 | 12 |
| 6. (3O,1AF) | $\{o_1, o_2, o_3\}$ | 10 | 20 |
| 7. (3O,3SR) | $\{o_1, o_2, o_3\}$, $\{R_1, R_2, R_3\}$, $\{v_1, v_2\}$, $\{v_3, v_4\}$, $\{v_5, v_6\}$ | 6 | 6 |
| 8. (3O,1AR) | $\{o_1, o_2, o_3\}$ | 10 | 20 |
| 9. (3O,1AR) | $\{o_1, o_2\}$ | 21 | 42 |

*Note.* One or more transformations are listed for each domain, and two concepts are instances of the same type if they are equivalent up to one or more of these transformations. The transformations are represented as sets, and permuting the entities within any set is a valid transformation. The second-to-last column shows the number of four-item types if exchanging the markers of concept membership (in and out) is a valid transformation. The final column shows the number of four-item types if in and out cannot be exchanged.

(*Appendices continue*)

Domain (3O,1SF) in Table 2 is very similar to domain (3O,1AF), and the only difference is whether the single feature in the domain is substitutive or additive. Table A1 indicates, however, that this difference leads to different numbers of concept types within the two domains. The 10 types of size four for domain (3O,1AF) are shown in Figure 2d. The only difference between the types in Figures 2c and 2d is that type 8 in Figure 2c is broken into two types (8 and 9) in Figure 2d. In domain (3O,1SF), exchanging the roles of $v_1$ and $v_2$ reveals that types 8 and 9 in Figure 2d are equivalent. When the single feature is additive, however, no corresponding transformation is possible, because the absence of a feature is qualitatively different from its presence. Of the eight domains in Table 2, the four that include additive features or relations each yield the 10 concept types shown in Figure 2d.

Although Shepard et al. (1961) pointed out that the concept types within a given domain can be computed by identifying structure-preserving transformations, their own work is not always consistent with this insight. In their second experiment, they compared concept learning across three domains that correspond to (1O,3SF), (3O,3SF), and (3O,1SF) in Table 2. Table A1 suggests that the first two domains both yield the six SHJ types, which means that it is sensible to compare how these types are learned within these domains. Domain (3O,1SF), however, is qualitatively different and includes nine types rather than six. Shepard et al. did not acknowledge this difference and considered the same SHJ types for all three domains in their experiment. They concluded that all three domains lead to similar behavioral results, but this conclusion deserves further examination. Figure 2c suggests, for example, that SHJ type IV corresponds to two qualitatively different types in domain (3O,1SF). As mentioned earlier, type 5 in (3O,1SF) can be described as "items that include two or more objects with value $v_2$ on $F$," but type 6 admits no similarly intuitive description. Experiment 1 in this article did not consider (3O,1SF), but the results for domain (3O,1AF) suggest that domain (3O,1SF) is likely to lead to different patterns of learning than domains (1O,3SF) and (3O,3SF).

Subsequent researchers have also failed to recognize that the eight domains in Table 2 are not truly isomorphic. One relevant study explored a version of domain (3O,1AR) where each item is a miniature social network that includes three people and specifies whether or not an undirected relation (e.g., friendship) exists between each pair of people (Crockett, 1982). Table A1 suggests that there are 10 types of size four in domain (3O,1AR), but the study of Crockett (1982) takes the six SHJ types for granted. As described in the main text, a careful characterization of the conceptual universe undermines the conclusions that Crockett drew from his results.

A second study carried out by social psychologists focuses on a version of domain (3O,1AR) in Table 2 where two of the objects are people, and the third is a social issue (Cottrell, 1975). For example, each network might indicate whether a positive attitude exists between two individuals $o_1$ and $o_2$, and whether these individuals view issue $o_3$ in a positive light. Cottrell (1975) has suggested that the six SHJ types capture all concepts of size four that exist within the domain, but Table A1 provides a very different perspective. Row 8 suggests that domain (3O,1AR) produces 10 distinct types when the labels of the three objects can be permuted. When the first two objects are people, and the third represents an issue, only the labels of the first two objects can be permuted, and row 9 of the table indicates that 21 types can be distinguished under these conditions. Any study organized around the six SHJ types is therefore unlikely to provide an adequate account of learning in this domain.

(*Appendices continue*)

## Appendix B

### Computing Minimal Descriptions

Complexity values for all languages in this article were computed by generating rules in order of increasing complexity until rules for each concept of interest had been found. The generation procedure for each language included an enumeration phase and a combination phase. The enumeration phase generated a set of rules according to criteria described in the next paragraph. Each rule has an extension that specifies which items in the domain are consistent with the rule. Given the extensions of all rules generated during the enumeration phase, the combination phase considered all possible ways to combine these extensions using conjunctions or disjunctions. The procedure terminated once extensions corresponding to all of the concepts under consideration had been found. Although the number of possible rules grows rapidly as the complexity of these rules increases, the number of extensions is fixed and relatively small ($2^8$ for domains of size 8). The combination phase is tractable because rules with the same extension can be grouped into a single equivalence class.

The enumeration phase varied according to the language under consideration. When computing complexities with respect to propositional logic, the enumeration phase generated all rules that correspond to a single literal. Because all literals were included and all propositional rules can be constructed as combinations of these literals, the complexity values for the propositional model are exact.

When computing complexities for languages OQ, FQ, and OQ + FQ, the enumeration phase considered all rules that had at most $m$ quantifiers and that had a complexity value less than or equal to $n$. All analyses for Experiment 1 used $m = 2$ and $n = 4$. For example, the enumeration phase for OQ did not include the rule $\exists_x \exists_y \exists_z\, F'_x F_y F_z$ (too many quantifiers) or the rule $\forall_x \exists_y F_y F_x + F'_x F'_y + G_y G_x$ (complexity too high). The same values of $m$ and $n$ were used when computing complexities for domain ($6O \times 1O$) in Experiment 2, but $m$ and $n$ were both set to 3 when computing complexities for domain ($3O \times 2O$) in Experiment 2.

There are algorithms other than the one just described that can be used to compute minimal descriptions with respect to different languages. Several researchers have developed algorithms for computing Boolean complexity (Lafond, Lacouture, & Mineau, 2007; Mathy & Bradmetz, 2004; Vigo, 2006), and researchers who work on inductive logic programming (Muggleton & De Raedt, 1994) have developed general heuristics for finding the shortest logical description of a data set. It is possible that some of the same heuristics can be applied to the languages considered in this article, but I am aware of no previous attempts to compute minimal descriptions with respect to languages like OQ.

(*Appendices continue*)

## Appendix C

## The Adjusted OQ Model

The OQ model accounts well for Feldman's (2000) data set only if adjusted to allow for two strategies that some participants may have used. When asked to learn a concept where the positive examples outnumber the negative examples, I assume that each participant will choose to encode the negative examples with probability $p$. When encoding a set of examples, I assume that with probability $q$, a participant will rely on a brute force strategy rather than identifying the shortest possible representation. If the brute force strategy is used, then the description length corresponds to the number of examples multiplied by the number of features possessed by each item. For any concept where the negative examples outnumber the positive examples, the predicted complexity is therefore

$$(1 - p)((1 - q)l^+ + qb^+) + p((1 - q)l^- + qb^-) \quad \text{(C1)}$$

where $l^+$ and $l^-$ are the minimum description lengths of the positive and negative examples, and $b^+$ and $b^-$ are the brute-force description lengths of the positive and negative examples. To avoid fitting numerical parameters, I set $p = q = 0.5$.

When participants are required to learn a concept where the negative examples outnumber the positive examples, Equation C1 is altered to capture the assumption that all participants choose to encode the positive examples. The predicted complexity is therefore

$$(1 - q)l^+ + qb^+. \quad \text{(C2)}$$

For most analyses in the article, the adjusted and unadjusted OQ models produce identical correlations with the human data. As described in the text, adjusting the OQ model is equivalent to adding a constant to each minimal description length whenever all concepts in a given family have the same size and the number of positive examples is smaller than or equal to the number of negative examples. Table C1 shows adjusted and unadjusted correlations for all data sets other than Feldman's (2000) data where the two approaches lead to different results. The table shows that the unadjusted and adjusted OQ models perform similarly in all four cases. Overall, then, the unadjusted OQ model accounts relatively well for all data sets except Feldman's data set, and the adjusted OQ model accounts well for all data sets considered.

Table C1
*Correlations Achieved by the Adjusted and Unadjusted OQ Models*

| Domain | Concept size | Unadjusted | Adjusted |
|---|---|---|---|
| (1O,2SF) | 3/6 | 0.99 | 0.99 |
| (1O,2SF) | 4/5 | 0.90 | 0.91 |
| (6O × 1O,1AR) | 4/5 | 0.88 | 0.79 |
| (3O × 2O,1AR) | 4/5 | 0.85 | 0.86 |

Equations C1 and C2 can also be used to adjust the mental models approach, but in this case, the description lengths $l^+$, $l^-$, $b^+$, and $b^-$ should be computed by counting the number of disjuncts rather than the number of literals. Results for the adjusted mental models approach on Feldman's (2000) data set are shown in the caption of Figure 6.

Received August 1, 2011
Revision received May 30, 2012
Accepted June 13, 2012 ■